Genetic Code Expansion in Mammalian Cells

Václav Beránek

MRC Laboratory of Molecular Biology

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Magdalene College

September 2018

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text

It does not exceed the prescribed word limit.

Václav Beránek

September 2018

Abstract

Genetic Code Expansion in Mammalian Cells

Václav Beránek

Proteins in nature are synthesized from a conservative set of 20 canonical amino acids, limiting the chemical space of biological systems. Over the last few decades, scientists have developed methods to expand the genetic code of living organisms, introducing new, non-canonical amino acids with diverse chemistries into proteins. These methods rely on engineering of the translational machinery often importing an aminoacyl-tRNA synthetase (aaRS)/tRNA pair that does not cross-react with the endogenous aaRS/tRNA pairs. This approach has allowed for co-translational incorporation of a variety of non-canonical amino acids, including amino acids for photocrosslinking, biophysical probes, amino acids bearing post-translational modifications or bio-orthogonal chemical handles for protein labelling and imaging. These designer amino acids allow researchers to probe, image and control protein function *in vivo* with great precision and minimal perturbation.

Cultured mammalian cells present an attractive model for studying human biology and disease. The first aaRS/tRNA pair used for genetic code expansion in mammalian cells was the *Escherichia coli* (*Ec*) tyrosyl-tRNA synthetase (TyrRS) and *Bacillus stearothermophilus* (*Bs*) tyrosyl-tRNA ($tRNA^{Tyr}$), however, this pair was quickly supplanted by *Methanosarcina barkeri* (*Mb*) and *Methanosarcina mazei* (*Mm*) pyrrolysyl-tRNA synthetase/tRNA pairs ($PylRS/tRNA^{Pyl}$), which are easily engineered to incorporate a wide variety of useful non-canonical amino acids. The developments in the genetic code expansion in mammalian cells have allowed researchers to study role of post-translational modifications, dissect signalling pathways, label and identify cell proteomes and identify protein-protein interactions directly *in vivo*.

This thesis presents several key advances in, and applications of, genetic code expansion in mammalian cells.

Chapter 1 introduces the relevant aspects of protein translation and summarises the progress in the field of genetic code expansion to date.

Chapter 2 describes genetic encoding of phosphoserine and its non-hydrolyzable analogue in mammalian cells. The engineered phosphoseryl-tRNA synthetase/tRNA pair ($SepRS^{v1.0}/tRNA^{v1.0}_{CUA}$) derived from *Methanococcus mariplaudis* and *Methanocaldococcus janaschii* (*Mj*) is shown to be orthogonal in mammalian cells. Subsequently, phosphoserine incorporation into a reporter protein is optimised by engineered translation elongation factor 1 alpha, engineered eukaryotic release factor 1 and metabolic engineering of the mammalian cell line. Overall this approach achieves an order of magnitude improvement in protein yield over unoptimised system. Incorporation of non-hydrolysable phosphonate analogue of phosphoserine in an engineered cell line is subsequently demonstrated and used for synthetic activation of a protein kinase.

Chapter 3 describes two methodological advances in the use of genetic code expansion for protein imaging. Firstly, we demonstrate the use of genetic code expansion for super-resolution microscopy. Secondly, we identify aberrantly extended endogenous proteins ending with the amber stop codon as a source of non-specific labelling. We then proceed to minimise this background by optimisation of the labelling protocol and use the resulting protocol for live-cell imaging of a recently discovered microprotein.

Chapter 4 experimentally demonstrates the orthogonality of a recently discovered $PylRS/tRNA^{Pyl}$ pair from *Methanomethylophilus alvus* (*Ma*) in mammalian cells. We further demonstrate that this pair is mutually orthogonal to the widely used *Mm*$PylRS/tRNA^{Pyl}$ pair, establishing a new, orthogonal pair for genetic code expansion in mammalian cells. The two pairs are used to site-specifically direct two distinct amino acids into a reporter protein.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Jason Chin, for his advice and guidance throughout my PhD. His wisdom and drive for excellence have taught me many lessons both inside and outside of the laboratory. My thanks also extends to my second advisor, Nick Barry for his generous support and many stimulating discussions.

I would further like to thank all of the current and former members of the Chin lab, whose advice and camaraderie have been instrumental to the completion of this thesis. Special thanks extends to my partners in crime Tao Uttamapinant and Roberto Zanchi who have patiently taught me everything I know about experimental biology. Our discussions, both scientific and non-scientific, have made for some of the most enjoyable moments in my PhD. I would further like to thank my partner on the phosphoserine project - Christopher Reinkemeier. I could not have wished for a better master student to work with. My thanks extends also to Toke Hanson, Simon Brunner and Inja Radman who have been amazing colleagues and great friends. Lastly, I would like to thank all of the members of team genome whose fearless work on an ambitious project made science feel like a true adventure: Julius Fredens, Kaihang Wang, Daniel de la Torre, Louise Funke, Wes Robertson, Yonka Christova, Tiongsun Chia, Wolfgang Schmied, Daniel Dunkelmann, Andres Gonzales Llamazares and Thomas Elliott.

I would like to express my thanks to the entire Cambridge University Ice Hockey Club, a team of individuals that have taught me more about leadership, friendship and determination than anyone else. Special thanks belong to Kumaran Nathan, Oscar Wilsby, Michal Barabas, Jaason Geerts, Julien Gagnon and Spencer Brennan.

Lastly, I would like to thank my girlfriend Anna Stejskalova, a true partner, friend and a companion in my life and my parents Pavel Beranek and Eva Berankova for their unwavering support and the values they have instilled in me.

Contents

List of Figures

Publications

The following publications include work discussed in this dissertation:

Maywood, E.S., Elliott, T.S., Patton, A.P., Krogager, T.P., Chesham, J.E., Ernst, R.J., Beránek, V. Brancaccio, M.m Chin J.W., Hastings, M.H., 2018. **Tranlsational switching of Cry1 protein expression confers reversible control of circadian behaviour in arrhythmic Cry-deficient mice.** *Proceedings of the National Academy of Sciences*, 115 (52), e12388-12397.

Beránek, V., Willis J.C.W. and Chin, J.W., 2018. **An evolved M. alvus pyrrolysyl-tRNA synthetase/tRNA pair is highly active and orthogonal in mammalian cells.** *Biochemistry*, 58 (5), p. 387-390.

Baumdick M., Gelléri M., Uttamapinant C., Beránek V., Chin J.W., Bastiaens P.I.H., 2018. **A conformational sensor based on genetic code expansion reveals an autocatalytic component in EGFR activation.** *Nature Communications*, 9 (3847)

Beránek, V., Reinkemeier, C.D., Zhang, M.S., Liang, A.D., Kym, G. and Chin, J.W., 2018. **Genetically Encoded Protein Phosphorylation in Mammalian Cells.** *Cell Chemical Biology*, 25(9), pp. 1067-1074.

Krogager, T.P., Ernst, R.J., Elliott, T.S., Calo, L., Beránek, V., Ciabatti, E., Spillantini, M.G., Tripodi, M., Hastings, M.H. and Chin, J.W., 2018. **Labelling and identifying cell-specific proteomes in the mouse brain.** *Nature Biotechnology*, *36*(2), p.156-159.

Ernst, R.J., Krogager, T.P., Maywood, E.S., Zanchi, R., Beránek, V., Elliott, T.S., Barry, N.P., Hastings, M.H. and Chin, J.W., 2016. **Genetic code expansion in the mouse brain.** *Nature Chemical Biology*, *12*(10), p.776-778.

Uttamapinant, C., Howe, J.D., Lang, K., Beránek, V., Davis, L., Mahesh, M., Barry, N.P. and Chin, J.W., 2015. **Genetic code expansion enables live-cell and super-resolution imaging of site-specifically labeled cellular proteins.** *Journal of the American Chemical Society*, *137*(14), p.4602-4605.

Chapter 1 Introduction

1.1 Genetic Code

The proteins of all living organisms known to date are synthesised from a limited set of building blocks - the 20 canonical amino acids. The field of genetic code expansion encompasses the efforts to create living organisms that incorporate additional amino acids, extending the chemical space of biology.

According to the central dogma of biology, the information defining the amino acid sequence of proteins is encoded in deoxyribonucleic acid (DNA). DNA is a chain molecule comprised of four nucleotides (adenine (A), thymine (T), guanine (G) and cytosine (C)) linked by a covalent bond in an alternating sugar-phosphate backbone (Watson and Crick, 1953). The DNA nucleotide sequence is copied into messenger RNA (mRNA) (also comprised of four nucleotides with thymine replaced by uracil (U)) based on Watson-Crick base pair complementarity in the process of transcription. In the last step of the information flow, the mRNA sequence is used as a template for protein synthesis by the translation machinery.

The set of rules that living organisms use to convert information from mRNA to proteins is know as the genetic code. The genetic code is constrained by the nature of the two polymers involved: while mRNA is a polymer chain of only four nucleotides, it encodes a polypeptide chain comprised of at least 20 various amino acids. Therefore, more than two nucleotides are needed to encode a single amino acid.

Deciphering of the genetic code – the set of rules how four nucleotides code for 20 various amino acids – has been one of the major achievements in molecular biology. The efforts started in 1960 with the work of Tsugita and Fraenkel-Conrat who demonstrated that the genetic code is non-overlapping (Tsugita and Fraenkel-Conrat, 1960; Wittmann, 1961). Crick et al. subsequently established, through mutagenesis of the bacteriophage T4, that the genetic code uses non-punctuated triplets, or codons (Crick et al., 1961). In a set of experiments, Crick et al.

demonstrated that only triple insertions (rather than single or double) lead to wild type phenotype of the bacteriophage.

In 1961, Nirenberg and Mattahei used cell-free protein synthesis experiment to show that polyuridine RNA leads to synthesis of polyphenylalanine (Nirenberg and Matthaei, 1961). In combination with the known triplet nature of the genetic code, the first decoded codon was therefore the triplet UUU coding for phenylalanine. Using analogous cell-free translation experiments combined with new approaches to synthesize polynucleotides of defined sequences, the remainder of the genetic code was deciphered over the following few years (Khorana et al., 1966). Because the 64 possible triplet codons ($4^3 = 64$) code only 20 different amino acids, the genetic code is redundant, i.e. several triplets can encode the same amino acid. In addition, three stop codons are present in the genetic code: UAA, UGA and UAG, which signal termination of translation.

The genetic code is strikingly conserved across all branches of life. According to the 'frozen accident' hypothesis, the genetic code, once assigned, became too costly or deleterious to change (Crick, 1968). However, relatively shortly after this hypothesis, some exceptions to the genetic code in different organisms (Lozupone et al., 2001), and in mitochondrial genomes have been found (reviewed in (Knight et al., 2001)). The growing list of alternative genetic codes has recently been expanded by a report of stochastic decoding of a codon in yeast *Ascoidea asiatica* (Mühlhausen et al., 2018). Despite these notable exceptions the majority of nuclear genetic codes conform to the standard, near-universal genetic code (**Figure 1.1**).

	AGA									UUA					AGC				
	AGG									UUG					AGU				
GCA	CGA						GGA			CUA				CCA	UCA	ACA			GUA	
GCC	CGC						GGC		AUA	CUC				CCC	UCC	ACC			GUC	UAA
GCG	CGG	GAC	AAC	UGC	GAA	CAA	GGG	CAC	AUC	CUG	AAA		UUC	CCG	UCG	ACG		UAC	GUG	UAG
GCU	CGU	GAU	AAU	UGU	GAG	CAG	GGU	CAU	AUU	CUU	AAG	AUG	UUU	CCU	UCU	ACU	UGG	UAU	GUU	UGA
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	stop
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	

Figure 1.1 The near-universal genetic code. Taken from (adapted from Alberts et al., 2002).

1.1.1 Fidelity of the genetic code: tRNAs and aminoacyl-tRNA synthetases

In the process of translation, the codon on the mRNA is paired with the particular amino acid inserted into the growing polypeptide chain via an adaptor molecule – the transfer RNA (tRNA). The tRNA is a short RNA molecule (typically less than 90 nucleotides long) that folds into a characteristic cloverleaf structure based on intra-molecular Watson-Crick base pairs (Sharp et al., 1985). The secondary structure forms three hairpin arms (D arm, anticodon arm and TΨC arm) and the acceptor (or aminoacyl) stem, which contains both the 5'-terminal phosphate group and the 3'-terminal CCA group (**Figure 1.2**).

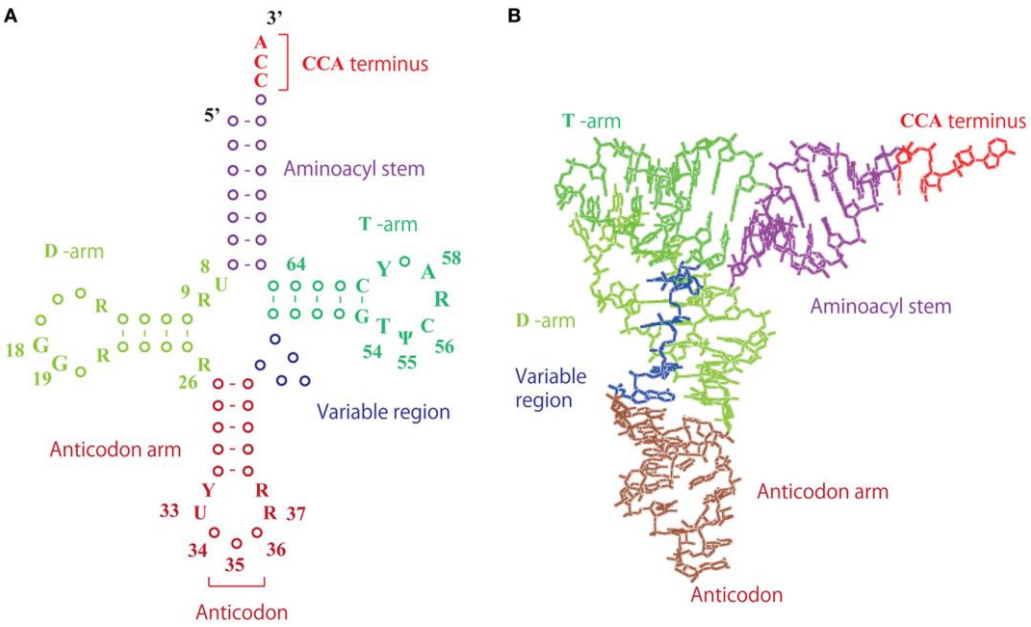

Figure 1.2.tRNA Structure. (**a**) The cloverleaf schematic showing Watson-Crick base pairing. (**b**) Tertiary structure of a typical tRNA molecule. (from (Hori, 2014)).

For translation, the CCA group on the acceptor stem is aminoacylated by an aminoacyl-tRNA synthetase (aaRS) in two steps: 1. ATP dependent adenylation of the cognate amino acid, forming aminoacyl-AMP intermediate, and 2. linkage of the α-carboxylate to the tRNA via either the 2'-OH or 3'-OH of the tRNA 3'-ribose (Arnez and Moras, 1997; Ibba and Söll, 2000).

Once aminoacylated, the tRNA anti-codon is matched with the triplet codon of the mRNA inside the ribosome.

The fidelity of the genetic code is maintained by strict specificity of the aminoacyl-tRNA synthetase (aaRS)/tRNA pairs: a unique tRNA has to be aminoacylated by only the corresponding amino acid. The high fidelity (with the error rate as low as 10^{-5} (Yamane and Hopfield, 1977)) is maintained by structural discrimination of both the amino acid substrate as well as the corresponding tRNA. On the tRNA, number of discriminating features, or identity elements, are present to ensure the stringency of the tRNA-synthetase selectivity. Most commonly these include the acceptor stem, the anticodon loop and the discriminator base (N73) (Ibba and Söll, 2000). In addition, crucial interactions can occur with specifically modified bases (Björk, 1995; Senger et al., 1997), in the variable arm (Hartz et al., 1990) or the D arm (Rasmussen et al., 2009). At the amino acid level, aaRSs have evolved several distinct features to prevent misacylation by the wrong amino acid. While some amino acids are sufficiently chemically distinct, for example cysteine, which is recognised via interaction with a Zn ion in the active site (Fersht and Dingwall, 1979; Newberry et al., 2002), some amino acids are very similar, for example tyrosine and phenylalanine, which differ by only a single hydroxyl group, or isoleucine and valine. To discriminate between the structurally similar amino acids, aaRSs use an additional editing domain, which proofreads the correct aminoacylation and hydrolyses the misactivated aminoacyl-adenylates (aa-AMPs) (reviewed in (Yadavalli and Ibba, 2012)). For example, a mutation in the editing domain of the isoleucyl-tRNA synthetase leads to misacylation rato of 1 in 150 (Fersht, 1977), while the wild type synthetase maintains accuracy of approximately 1 in 3000 (Loftfield and Vanderjagt, 1972).

1.2 Translation

The process of protein biosynthesis (translation) is mediated by ribosomes, large ribonucleoproteins composed of two subunits with many key components conserved across all domains of life. The ribosomes in prokaryotes consist of small (30S) and large (50S) subunit, which combine into the 70S ribosome, a 2.5-megadalton (MDa) structure. The eukaryotic ribosome is similarly composed of small (40S) and large (60S) subunits that form the 80S

ribosome with size between 3.5 and 4.5 MDa (Melnikov et al., 2012). The small subunit is responsible tor binding of the mRNA and – at its interface with the large subunit – forms the decoding centre where codon and anti-codon pairing occurs. The large subunit of the ribosome contains three tRNA binding sites on its interface side: the A site (which binds aminoacyl-tRNA), the P site (which binds peptidyl-tRNA) and the E site (which binds free tRNA before exit). In addition, the large subunit contains the peptidyl transferase centre where the catalysis of peptide bond occurs.

Canonically, the first translated codon is AUG, which is recognised by a dedicated initiator methionyl-tRNA (Met-tRNAi). In bacteria, the methionyl-tRNA is first formylated by formyltransferase resulting in formyl-methionyl-tRNA (fMet-tRNAi).

1.2.1 Prokaryotic Translation

In *E. coli*, translation starts through binding of the Shine-Dalgarno (SD) sequence on the mRNA to the complementary anti-SD sequence in the 16s ribosomal RNA (rRNA). Since the SD sequence is positioned just upstream of the start codon AUG, the 30S ribosomal subunit is recruited directly to the initiation region of the mRNA. In contrast to *E. coli*, other Gram-negative phyla do not use SD recognition to initiate translation (Accetto and Avgustin, 2011), and leaderless mRNAs and mRNAs lacking SD are very common in bacterial genomes (Chang et al., 2006), therefore the conservation of this mechanism throughout the prokaryotic domains remains unclear.

Three initiation factors are involved in translation initiation in *E. coli*: prokaryotic initiation factors 1,2, and 3 (IF1, IF2, and IF3). The IF3 and IF2 are the first proteins to bind to the 30S subunit, with subsequent binding of IF1, which enables the recruitment of the fMet-tRNAi and stabilization of the 30S initiation complex (Masuda et al., 2012; Milón et al., 2012). Interestingly, the interaction of 30S subunit with mRNA is independent of binding of the IF2 and IF3 and can happen at any point before the binding of IF1 (Milón et al., 2012). Once formed, the 30S initiation complex with fMet-tRNAi is quickly bound by the 50S ribosomal subunit, forming the 70S initiation complex. The IF2 subsequently dissociates, allowing for

binding of the elongation factor thermally unstable (EF-Tu) and start of the elongation phase (Gualerzi and Pon, 2015).

In elongation, aminoacylated tRNAs are delivered to the A-site of the ribosome by the EF-Tu in a ternary complex EF-Tu-GTP-aminoacyl-tRNA. The affinities of EF-Tu-GTP for the various aminoacylated tRNAs are very high (K_d ~ nM) (Louie and Jurnak, 1985) and, interestingly, span only 1 order of magnitude due to the thermodynamic compensation (LaRiviere et al., 2001). The thermodynamic compensation posits that the combined contribution of the tRNA and its cognate amino acid to EF-Tu binding add to the same level across all aminoacyl-tRNA species. This mechanism ensures appropriate level of different ainoacyl-tRNA-EF-Tu complexes available for protein synthesis and helps exclude misaminoacylated tRNAs, which are either bound too weakly or too strongly by the EF-Tu, from translation (Schrader et al., 2011; Schrader and Uhlenbeck, 2011). Once the correct anticodon is matched with the current codon in the A-site, EF-Tu in combination with the small ribosomal subunit triggers first hydrolysis of GTP and subsequently conformational switch of EF-Tu and its release (Stark et al., 1997). The free aminoacyl-tRNA can fully accommodate the A-site and formation of the peptide bond and simultaneous decaylation of the tRNA in the P-site can occur. After the bond formation, the ribosome translocates by a codon-sized step (approx. 10 Å) in a process catalysed by elongation factor G (EF-G), transferring the newly formed peptidyl-tRNA from A-site to P-site and the deacetylated tRNA from P-site to E-site (Achenbach and Nierhaus, 2015). This process completes one elongation step after which additional tRNAs are shuttled to the ribosome via the elongation factor until the ribosome encounters the first stop codon.

The translation is terminated when one of the three stop codons (UAA, UGA or UAG) is positioned in the A-site. Unlike codons responsible for initiation and elongation, the stop codons are recognised by proteins called class I release factors (Capecchi, 1967; Vogel et al., 1969). The two proteins responsible for termination in bacteria are release factor 1 and 2 (RF1 and RF2). Interestingly, each of the two release factors responds to specific, yet overlapping set of stop codons: EF1 terminates at UAG and UAA while EF2 terminates at UGA and UAA (Ito et al., 1996; Nakamura et al., 1995). Upon binding of one of the release factors, the ester

bond joining the peptidyl moiety and terminal nucleotide of the tRNA in the A site is hydrolysed and the newly synthesized peptide is released (Korostelev, 2011).

1.2.2 Eukaryotic translation

Translation in eukaryotes is significantly more complex and involves a number of factors. The process starts with the preinitiation complex (PIC) binding to 5' end of mRNA in a reaction promoted by initiation factors eIF1, eIF1A, eIF5 and eIF3. The PIC is composed of the small (40S) ribosomal subunit and the initiator methionyl-tRNA (Met-tRNAi) anchored together by GTP-bound form of eIF2. Binding of the PIC to the mRNA is facilitated by several additional eukaryotic initiation factors that recognise and bind the 5'-cap structure of the mRNA (eIF4E, eIF4G and eIF4A). After binding, the initiation complex scans the mRNA base-by-base until the first AUG is encountered and matched with the Met-tRNAi in the P decoding site of the ribosome (the sequence preceding the start codon is also involved and should conform to the Kozak consensus sequence (Kozak, 1987)). Matching of the AUG codon triggers hydrolysis of GTP in eIF2-GTP-Met-tRNAi complex, which releases eIF2-GDP and other eIFs and leads to recruitment of the large (60S) ribosomal subunit, forming the 80S initiation complex (Pestova et al., 2007).

Similarly to prokaryotic elongation, aminoacyl-tRNAs are delivered to the 80S ribosome via the alpha subunit of the eukaryotic elongation factor 1 (eEF1α) in eEF1α-GTP-aminacyl-tRNA complex (Carvalho et al., 1984). Once correct match based on Watson-Crick base pairing between the mRNA codon in the A-site and the tRNA anticodon occurs, conformational change in the ribosome triggers GTP hydrolysis and dissociation of the eEF1α-GDP complex (Taylor et al., 2007). The sequences of eEF1α and EF-Tu are relatively homologous and (due to only limited structural data available (Andersen et al., 2000; Crepin et al., 2014)) most of the current knowledge relies on studies of EF-Tu.

In eukaryotes, unlike in bacteria, a single release factor is responsible for recognition of all three stop codons (Brown et al., 2015; Frolova et al., 1994). The translation is terminated when the eukaryotic release factor 1 (eRF1) and the co-factor eukaryotic release factor 3 (eRF3) form eRF1-eRF3-GTP complex and bind to the stop codon in ribosomal A-site (Muhs et al., 2015;

Taylor et al., 2012). Upon binding, GTP hydrolysis leads to dissociation of eRF3, bringing the GGQ motif of eRF1 close to the ester bond between the tRNA and the polypeptide leading to its hydrolysis triggering translation termination and subsequent dissociation of the ribosome through action of ABCE1 ATPase (Pisarev et al., 2010; Shoemaker and Green, 2011).

1.3 Post-translational modifications

The chemical space of proteins is expanded beyond the 20 canonical amino acids through a number of post-translational modifications (Walsh et al., 2005). Post-translational modifications are generally covalent, reversible, enzyme-mediated attachments of chemical groups to the side chains of the amino acids; however, they can also include proteolytic cleavage of functional parts of the protein, fluorescent protein maturation, or protein splicing. Currently, there are over 300 types of post-translational modifications known (Walsh, 2006).

The major post-translational modifications include phosphorylation, glycosylation, ubiquitination, methylation, acetylation, alkylation, oxidation and less frequently succinylation, SUMOylation and citrullination (**Figure 1.3**). Collectively, over 260,000 of post-translationally modified sites have been experimentally discovered (www.phosphosite.org, (Doll and Burlingame, 2014)), with many playing critical role in both physiological (Deribe et al., 2010; Hsu et al., 2011; Humphrey et al., 2015) and pathological processes (Blume-Jensen and Hunter, 2001; Zhang et al., 2009).

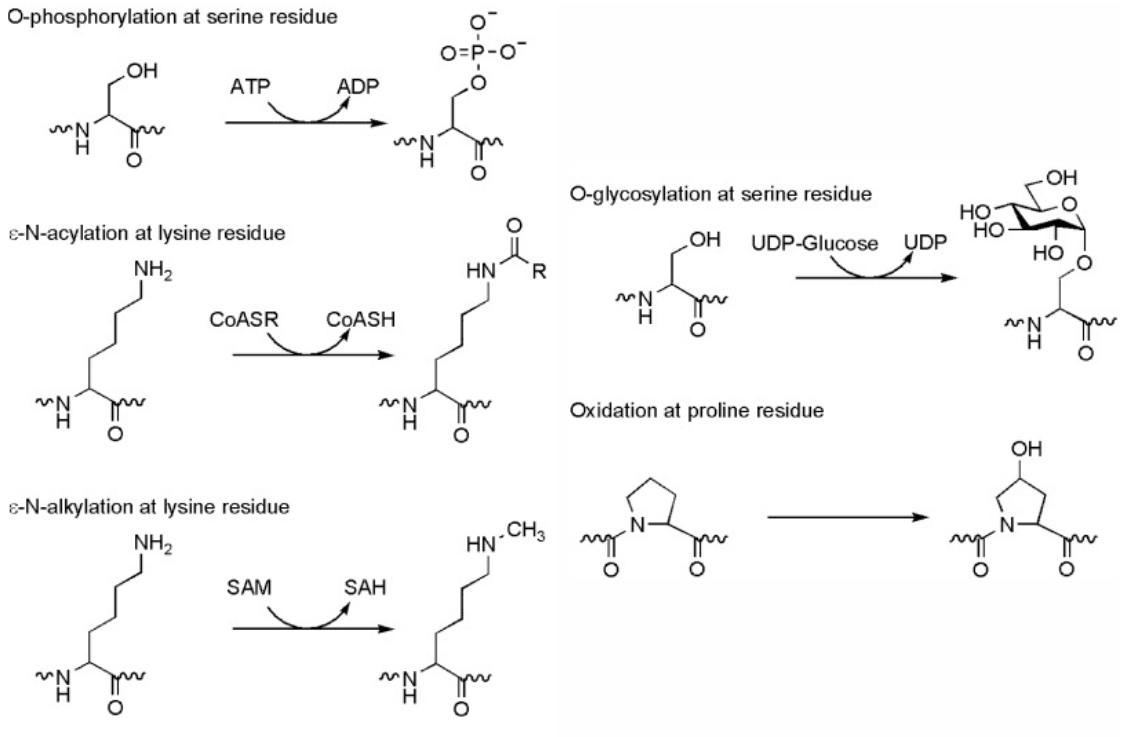

Figure 1.3. Major types of post-translational modifications. The most common post-translational modifications include pohosphorylation, acylation, alkylation, glycosylation and oxidation. Figure adapted from (Walsh et al., 2006).

1.3.1 Lysine acetylation

Acetylation was first discovered together with methylation through *in vitro* analysis of histones isolated from calf thymus (Allfrey et al., 1964) and has been postulated to regulate DNA transcription. The first enzyme responsible for installation of acetyl group from coenzyme A has been identified in 1996 in *Tetrahymena* (Brownell et al., 1996). Originally named histone acetyltransransferases, the enzymes were found to acylate wide variety of substrates and therefore subsequently renamed to lysine acetyltransferases (Allis et al., 2007; Drazic et al., 2016).

1.3.2 Phosphorylation

Perhaps the most studied post-translational modification is phosphorylation – the attachment of the phosphoryl group to the sidechain of one of the amino acids (Cohen, 2002). Currently, tens of thousands phosphosites have been experimentally reported in cultured cells (Humphrey et al., 2013; Lundby et al., 2012) and it is estimated that up to 1 million phosphosites concurrently exist in the mammalian proteome (Boersema et al., 2010; Lemeer and Heck, 2009). The most commonly phosphorylated amino acids in eukaryotes are serine, threonine and tyrosine (Olsen et al., 2006), and less commonly histidine (Steeg et al., 2003), with the distribution of frequency corresponding to 88:11:1 for phosphorylated serine, threonine and tyrosine, respectively (Olsen et al., 2006; Villén et al., 2007). Despite the abundance of phosphosites reported, only fraction of them has an identified physiological function. Not surprisingly, functionally relevant phosphosites tend to be evolutionarily conserved, suggesting a way to map out useful directions of research (Levy et al., 2012).

Phosphorylation can act as an activating modification, increasing or decreasing the protein's biological activity through the addition of negative charge to the modified amino acid. This observation has led to the intriguing hypothesis that key aspartic and glutamic acids in enzymes may have been substituted with phosphorylated amino acids to allow for reversible control over their activity (Pearlman et al., 2011). Alternatively, protein phosphorylation can stabilise proteins (Chehab et al., 1999), mark them for degradation (Yada et al., 2004), disrupt or mediate protein-protein interactions (Pawson and Nash, 2003) or facilitate transport of proteins between cellular compartments (Jans and Hubner, 1996).

1.4 Natural expansion of the genetic code: selenocysteine and pyrrolysine

When first deciphered, the genetic code specified the decoding rules for the 20 canonical amino acids (Khorana et al., 1966). The chemical diversity of proteins and as discussed in the previous section amino acids can be further extended through post-translational modifications; however, in rare occasions living organisms use entirely new amino acids (Ambrogelly et al., 2007).

1.4.1 Selenocysteine

Selenium was first discovered to be essential for activity of formate dehydrogenase in *E. coli* in 1954 (Pinsent, 1954) and the amino acid selenocysteine (Sec) was first identified in 1976 as the selenium moiety in glycine reductase in 1976 (Cone et al., 1976).

Surprisingly, the codon corresponding to the position of selenocysteine was found to be the TGA stop codon in both *E. coli* formate dehydrogenase (Zinoni et al., 1986) and mouse glutathione peroxidase (Chambers et al., 1986), suggesting another exception to the frozen genetic code hypothesis by Crick et al. The subsequent discovery of a designated tRNA responsible for directing selenocysteine and recognizing the UGA stop codon in *E. coli* (Leinfelder et al., 1988) implied co-translational incorporation of the amino acid and led to the designation of selenocysteine as the 21st amino acid (Söil, 1988).

In bacteria, the selenocysteinyl-tRNA (Sec-tRNA^{Sec}) is synthesized through an intermediate step where the tRNA^{Sec} is first aminoacylated with serine creating Ser-tRNA^{Sec}, which is then converted to Sec-tRNA^{Sec} through the action of L-seryl-tRNA^{Sec} selenium transferase (SelA) (Forchhammer, 1991; Böck et al., 2005). In archea and eukaryotes, the Ser-tRNA^{Sec} is first converted to phosphoseryl-tRNA^{Sec} via the action of *O*-phosphoseryl-tRNA^{Sec} kinase (PSTK) and the resulting phosphoseryl moiety is subsequently converted to selenocysteine by Sep-tRNA:Sec-tRNA synthase (SepSecS) (Yuan et al., 2006).

To discriminate between the UGA codons intended for incorporation of selenocysteine and UGA codons that signal translation termination, the cell uses a selenocysteine insertion sequence (SECIS) in the mRNA. First discovered in *E. coli* (Forchhammer, 1989) and subsequently in mammals (Berry et al., 1991) the SECIS stem-loop structure is positioned directly downstream (3' side) from the UGA codon in bacteria and in the 3' untranslated region in eukaryotes and archea. SECIS is bound by SELB protein in bacteria (Baron et al., 1993) and SECIS-binding protein 2 (SBP2) in mammals (Copeland et al., 2000; Fletcher et al., 2001). These proteins are responsible for recruiting a specialised elongation factor SELB in bacteria (Forchhammer et al., 1989) and mSelB in eukaryotes (Fagegaltier et al., 2000; Tujebajeva et al., 2000) that delivers the charged Sec-tRNA^{Sec} into the ribosome (Allmang and Krol, 2006).

Although most species do not use selenocysteine, it can be found throughout the three domains of life. It is virtually absent in higher plants and fungi with some minor exceptions (Hatfield et al., 1992; Obata and Shiraiwa, 2005). Similarly, relatively narrow range (up to quarter) of sequenced prokaryotes use selenocysteine (Kryukov and Gladyshev, 2004; Zhang et al., 2006). In humans, 25 genes code for total of 46 selenocysteine residues (Kryukov et al., 2003) mostly in active sites of peroxidases (Hondal et al., 2013) and reductases (Achilli et al., 2015). Of these, 24 are conserved in rodents (Kryukov et al., 2003). The function of selenoproteins has been summarised in review by Reeves and Hoffman (Reeves and Hoffmann, 2009).

1.4.2 Pyrrolysine

Pyrrolysine (Pyl) is the second naturally incorporated amino acid, outside the canonical set of 20. Pyrrolysine is a lysine with a pyrroline-ring attached to the ε–amino group and was first discovered in the methanogen *Methanosarcina barkeri* based on structure of the MMA methyltransferase (MtmB) that has an in-frame UAG codon (Hao et al., 2002; Srinivasan et al., 2002). In the *Methanosarcina* species, pyrrolysine is a critical residue in the enzymes called methylamine methyltrasnferases that catalyse formation of methane from methylamines (Krzycki, 2004). Outside of methylamine methyltransferases and transposases (Zhang et al., 2005), pyrrolysine was also found to be incorporated in $tRNA^{His}$ guanylyltransferase (Thg1), an enzyme that is responsible for tRNA processing, however, pyrrolysine does not participate in the activity of that particular enzyme and can be substituted by tryptophan (Heinemann et al., 2009).

Initially it was suspected that, like in the case of selenocysteine, pyrrolysyl-$tRNA^{Pyl}$ (Pyl-$tRNA^{Pyl}$) is formed enzymatically from Lysyl-$tRNA^{Pyl}$ (Srinivasan et al., 2002), however, it was subsequently confirmed that pyrrolysine is charged directly onto the $tRNA^{Pyl}$ through the action of pyrrolysyl-tRNA synthetase (PylRS) both *in vitro* and *in vivo* (Blight et al., 2004; Polycarpo et al., 2004) and that the amino acid is synthesized by enzymes encoded in the *pylBCD* genes (Gaston et al., 2011; Longstaff et al., 2007).

Unlike selenocysteine, pyrrolysine has been discovered only in a very narrow range of species from the archea domain, mainly in the *Methanosarcinaceae* family, and several species in the

bacterial domain (Prat et al., 2012). More recently, a new order of methanogens has been identified (Borrel et al., 2013), including three species with pyrrolysine systems: *Methanomethylophilus alvus* (*M. alvus*), *Methanomassiliicoccus intestinalis* (*M. intestinalis*) and *Methanomassiliicoccus luminyensis* (*M. luminyensis*) (Borrel et al., 2014) (Willis and Chin, 2018).

Currently, the rules for UAG decoding in the species with the pyrrolysine system are unclear (Zhang et al., 2005). Initial reports suggested that a pyrrolysine insertion sequence (PYLIS) element is present 5-6 nucleotides downstream of the UAG codon, acting in a fashion similar to SECIS (Namy et al., 2004) (Ibba and Söll, 2000), however, a subsequent analysis did not confirm this hypothesis (Zhang et al., 2005). Interestingly, some genomes in Pyl-utilising archea show dramatically lower-than-expected usage of the UAG codon (< 5% of genes putatively terminate with UAG). In addition, readthrough of these stop codons would result in only low number of gene overlaps. Thereby lack of UAG candidates for the stop signal could lead to the hypothesis that UAG codons in Pyl containing species do not result in translation termination. Contrary to this hypothesis, the UAG codon is relatively frequent in the bacterium *D. hafniense*., suggesting that ambiguous decoding of the UAG codon must be present (Zhang et al., 2005). The experimental evidence indicates that no special mRNA element is required, as transformation of the *pylS* and *pylT* genes into *E. coli* grown in presence of pyrrolysine analog leads to readthrough of an in-frame UAG codon inside a reporter protein coding sequence (Polycarpo et al., 2006). Lastly, there is no known elongation factor associated with the pyrrolysine system and Lys-tRNA^{Pyl} binds EF-Tu effectively *in vitro* (Namy et al., 2004), suggesting that no other accessory protein is required for Pyl incorporation at a given UAG codon.

1.5 Artificial genetic code expansion

Expanding the set of amino acids used by living organism allows researchers to endow the organisms with new functionalities through designed chemical and physical properties of the introduced building blocks. These approaches have found many uses in probing, imaging and control of protein function in their native environment or production of modified proteins for

in vitro applications (Chin, 2017). Simultaneously, expanding the genetic code can allow researchers to harness the synthetic power of the cell's translation machinery to produce entirely new, genetically determined unnatural polymers. These two overlapping goals – encoding new amino acids into proteins in cells and animals and using the cells machinery to synthesize unnatural polymers – fundamentally rely on re-engineering of the translation machinery as well as their genetic code (Chin, 2017).

1.5.1 *In vitro* expansion of the genetic code

Traditionally, peptides including non-canonical or site-specifically modified amino acids are prepared using solid phase peptide synthesis (SPPS) (Merrifield, 1963), a revolutionary method that has been widely adopted for synthesis of small amounts (< 1 g) of short peptides (< 50 residues). The field was further advanced by invention of native chemical ligation (NCL) (Dawson et al., 1994), which allowed for total synthesis and study of longer proteins (reviewed in Bondalapati et al., 2016), of up to 203 (Torbeev and Kent, 2007) and 304 amino acids (Kumar et al., 2011). However, the yield of these methods as well as the length of possible products is limited and the amide bond synthesis has significant drawbacks including its expense and environmental impact (Pattabiraman and Bode, 2011).

An alternative relies on site-specific chemical modifications of unique amino acids within the canonical set. Unfortunately, reactions that proceed under native conditions (aqueous solution at moderate temperature and neutral pH) are quite limited (Durek and Becker, 2005). The most widely used approach relies on the unique characteristics of the sulfhydryl group in cysteine residues, which can be derivatized by maledimides or methanethisulfonates, however, these techniques are inherently limited by the complex nature of proteins, where multiple amino acids within the chain exist.

Yet another alternative to chemical protein synthesis or semi-synthesis is *in vitro* translation. In 1962 Benzer's group first showed that cysteinyl-$tRNA^{Cys}$ could be converted to alanyl-$tRNA^{Cys}$ using Ranley nickel and subsequently used in translation to incorporate alanine at a cysteine codon (Chapeville et al., 1962). A general method to aminoacylate any tRNA *in vitro* using T4 RNA ligase was subsequently developed (Hecht et al., 1978; Heckler et al., 1984) and

the chemically mis-aminoacylated tRNAs were shown to be translationally competent in an *in vitro* translation system (Baldini et al., 1988). This technique allowed researchers to obtain a tRNA aminoacylated with non-canonical amino acid and led to the first *in vitro* expansion of the genetic code. The second step to be solved was the absence of a codon to encode the non-canonical amino acid. Inspired by nature, which uses one of the stop codons for encoding selenocysteine and pyrrolysine (Ambrogelly et al., 2007), Noren et al. were in a landmark experiment able to site-specifically incorporate a non-canonical amino acid by first mutating the anticodon of $tRNA^{Phe}$ to decode a stop codon UAG and then chemically aminoacylating the $tRNA^{Phe}_{CUA}$ with a range of non-canonical amino acids. The aminoacylated $tRNA^{Phe}_{CUA}$ was used to *in vitro* translate a mRNA of the β-lactamase with an in-frame UAG codon (Noren et al., 1989). A similar experiment was performed subsequently, incorporating iodotyrosine via *in vitro* aminoacylated $tRNA^{Gly}_{CUA}$ in response to the UAG stop codon in a peptide of 16 amino acids (Bain et al., 1989). Furhter experiments demonstrating incorporations of a variety of non-canonical amino acids followed (reviewed in (Cornish et al., 1995)). Unfortunately, this approach suffers from significant drawback – the chemically aminoacylated tRNA is a limiting reagent and its preparation is not trivial, rendering the preparation of sufficient amount of site-specifically modified proteins difficult.

Despite the difficulty of preparation of significant amounts of aminoacylated *in vitro* tRNAs, this approach can be very useful for single-cell studies, where only limited amount of material is needed. This was first achieved by co-microinjection of amber suppression $tRNA_{CUA}$ aminoacylated with tyrosine derivatives and mRNA molecule coding for alpha subunit of the muscle type Ach receptor (AChR) into the *Xenopus* oocyte (Nowak et al., 1995). The mRNA was translated by the endogenous cell machinery and full-length alpha subunit was produced AChR with site-specific insertion of the given non-canonical amino acid. A similar approach using transfection of aminoacylated tRNAs into mammalian cells was also demonstrated (Köhrer et al., 2001).

Using microinjection or transfection of *in vitro* aminoacylated tRNAs, the resulting number of full-length proteins is approx. 1000-fold lower than the amount of injected aminoacylated tRNA, limiting the applicability of this method to study of proteins with detectable phenotype

with only a low number of molecules inside cell, such as channels, receptors and transporters (Dougherty and Van Arnam, 2014).

Notably, the *in vitro* aminoacylation approach does not require a synthetase capable of aminoacylating its cognate tRNA with a given amino acid and thereby providing more flexibility than the classical *in vivo* approaches described in the next section.

1.5.2 *In vivo* expansion of the genetic code

1.5.2.1 "Surrogate amino acids"

Initial attempts at encoding non-canonical amino acids *in vivo* made use of the endogenous aminoacyl-tRNA synthetase/tRNA (aaRS/tRNA) pairs and non-canonical amino acids analogous to the endogenous ones, also called "surrogate amino acids". In one of the early experiments Cowie and Cohen were able to completely replace methionine with selenomethionine (SeMet) by supplementing it to a methionine auxotrophic strain of *E. coli*., which produced active β-galactosidase incorporating SeMet instead of methionine (Cowie and Cohen, 1957). Many similar studies have followed, using vaious microorganisms, including *Staphylococcus Bacillus* or *Salmonella*, and eukaryotic cells, including *Saccharomyces*, *Tetrahymena* or immortalised cell lines such as *HeLa* (reviewed in (Richmond, 1962)), incorporating variety of surrogate or modified amino acids (reviewed in Wilson and Hatfield, 1984). The proteome wide incorporation of SeMet by the endogenous Methinoyl-tRNA synthetase has proven to be useful many years later for solving the phase problem in X-ray crystallography via multiwavelength anomalous diffraction (MAD) developed by Hendrickson (Hendrickson, 1991).

Due to the fact that most amino acids outside the 20 canonical are toxic to cells, an alternative protocol called selective pressure incorporation (SPI) (also designated the auxotroph method or media-shift method) was developed (Budisa et al., 1999). In this protocol, the target gene expression is kept silent during the growth phase of the cells and induced only after sufficient number of cells is present. The natural amino acid is then depleted from the media and the non-canonical substitute is added. The cells serve as a factory for the protein production without further growth (Budisa et al., 1999). This method allows for production of labelled proteins on

a relatively large scale for structural studies (e.g. (Bae et al., 2001; Sykes et al., 1974)) or production of therapeutics (Budisa et al., 2001).

In a first example of an engineered aaRS, Kas and Hennecke identified a point mutant of *E. coli* PheRS, which showed relaxed substrate specificity. This mutant was discovered when searching for Phe binding site and showed the ability to aminoacylate the $tRNA^{Phe}$ with *p*-chlorophenylalanine (Kast and Hennecke, 1991). Later work demonstrated that the Ala294Gly mutant reported by Kas and Hennecke also accepts *p*-chlorophenylalanine and *p*-bromophenylalanine as a substrate (Ibba et al., 1994)and that additional Thr251Gly mutation further allowed for incorporation of variety of additional amino acids, including *O*-acetyltyrosine, *p*-iodophenylalanine, p-cyanophenylalanine or *p*-azidophenylalanine (Datta et al., 2002; Kirshenbaum et al., 2002).

1.5.2.2 Amber suppression

The strategies using close structural homologs of endogenous amino acids in combination with auxotrophic strains suffer from significant limitations. Firstly, the proteins expressed contain stochastically both endogenous and non-canonical amino acids; secondly, the non-canonical amino acid is incorporated in all positions of all proteins. As a consequence, the health of the organism is compromised. In order to truly expand the genetic code and have site-specific incorporation of non-canonical amino acids, two conditions must be met:

1. tRNA must deliver the non-canonical amino acid in response to a unique codon and;
2. the tRNA must not be used by any of the endogenous aaRSs present in the cell.

To solve the first issue, researchers have taken inspiration from organisms with naturally expanded genetic code (Chapter 1.7). These organisms achieve incorporation of selenocysteine by suppression of one of the stop codons, specifically the UAG amber stop codon. The amber stop codon is the least frequent stop codon both in *E. coli* (314 UAG codons (Alff-Steinberger and Epstein, 1994; Blattner et al., 1997) and can be supressed by tRNAs with modified anticodon loops (Michaels et al., 1990), suggesting that amber suppression is compatible with the cell survival. The alternative to stop codons is the use of quandruplet codons, which are

used in some organisms (Riddle and Carbon, 1973; Yourno and Kohno, 1972) (the use of quadruplet codons is further discussed in section 1.6.2).

The stop codon suppression approach is combined with import of a new, orthogonal aminoacyl-tRNA synthetase/tRNA (aaRS/tRNA) pairs (Liu and Schultz, 2010) (**Figure 1.4**). In this approach, the exogenous tRNA is aminoacylated by its cognate, often engineered aaRS *in vivo* with the non-canonical amino acid either supplemented externally or made bioavailable through genetic engineering of the target organism (Liu and Schultz, 2010; Zhang et al., 2017) (**Figure 1.4**).

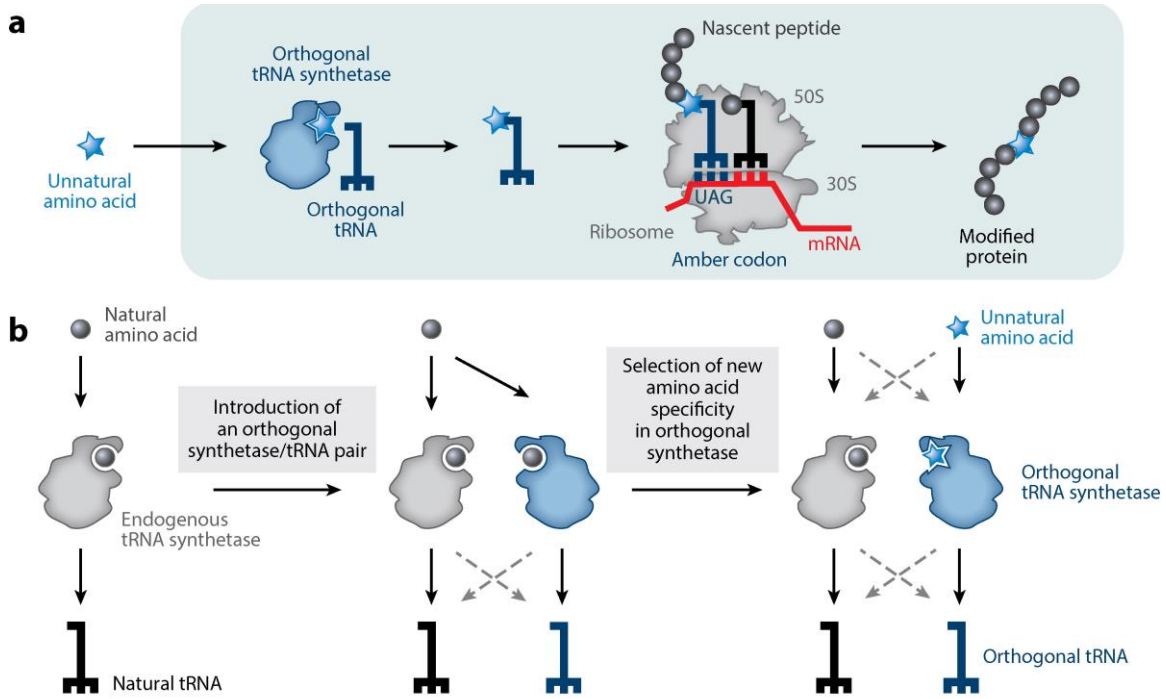

Figure 1.4. Amber supression. (a) Amber suppression relies on import of orthogonal tRNA synthetase/tRNA pair with CUA anticodon. (b) Typical evolution of a new, orthogonal aminoacyl-tRNA/tRNA pair which is orthogonal to the endogenous pairs. Figure adapted from (Chin, 2014).

The imported aaRS as well as the imported tRNA must be orthogonal to the endogenous tRNAs and aaRSs. The biological orthogonality is satisfied when the imported aaRS doesn't

aminoacylate any of the endogenous tRNAs and, at the same time, does not accept any of the endogenous amino acids as a substrate. Siumltaneously, the imported tRNA must not be aminoacylated by any of the endogenous aaRSs. The first step towards creation of an orthogonal aaRS/tRNA pair was taken by Liu and colleagues, who have engineered the GlnRS/$tRNA^{Gln}$ pair from *E. coli* based on the crystal structure of the aaRS:tRNA:ATP complex (Rould et al., 1989). In the first step, the $tRNA^{Gln}$ anticodon was changed from CUG to CUA and 3 sites on the $tRNA^{Gln}$, selected based on their contact with the GlnRS were mutated to abolish recognition, creating a new mutant designated O-tRNA. The GlnRS was then evolved using random point mutagenesis to aminoacylate the O-tRNA, however; only to a level well below the aminoacylation by endogenous GlnRS (Liu et al., 1997a; Liu et al., 1997b). Based on the observation that *E. coli* GlnRS does not charge yeast $tRNA^{Gln}$, due to additional domain present in the yeast homologue (Whelihan and Schimmel, 1997), Liu and Schultz developed a first orthogonal pair in *E. coli*. They were able to verify that 1. suppressor $tRNA_2^{Gln}{}_{CUA}$(A38) is not aminoacylated by any endogenous *E. coli* aaRSs *in vivo* using amber mutant of β-galactosidase and 2. yeast GlnRS aminoacylates the yeast $tRNA^{Gln}{}_{CUA}$ but not any of the endogenous *E. coli* tRNAs. However, the effort to evolve the yeast GlnRS to aminoacylate any non-canonical amino acids were not successful (Liu and Schultz, 1999).

1.5.3 *M. janaschii* tyrosyl-tRNA synthetase/tRNA

An alternative system based on tyrrosyl-tRNA synthetase (TyrRS) was developed by several groups. The TyrRS benefits from a lack of proofreading activity, which allows for an easier change in substrate specificity (Fersht et al., 1980; Jakubowski and Goldman, 1992). A first engineered TyrRS was reported by Hamano-Takaku and colleagues who used a random mutation library screen to isolate a Phe130Ser mutant which accepted 2-azatyrosine as a substrate (Hamano-Takaku et al., 2000). The first orthogonal aaRS/tRNA pair used in *E. coli* was based on *Methanococcus janaschii* TyrRS/tRNA, which does not aminoacylate endogenous *E. coli* tRNA to a high degree (Steer and Schimmel, 1999) but does aminoacylate its own amber suppressor $tRNA^{Gly}{}_{CUA}$ (Wang et al., 2000) likely due to absence of the C-terminal domain. In the first step, Wang et al. identified tRNA with decreased rate of aminoacylation by endogenous *E. coli* aaRSs based on a library screen with β-lactamase

selection. Subsequent rounds of selections from random mutagenesis library of the residues in the binding pocket of the *M. janaschii* TyrRS were then applied (Brick et al., 1989). The assay was based on survival conferred by chloramphenicol acetyltransferase (CAT) gene with an amber codon in a non-essential position Asp112 (Stemmer, 1994) in the presence of 1 mM *O*-methyl-*L*-tyrosine. Isolating the clones that grew in the presence but not in the absence of *O*-methyl-*L*-tyrosine led to identification of a mutant that supported expression of 2 mg/ml of *O*-methyl-*L*-tyrosine mutant of dihydrfolate reductase (DHFR) with over 95% purity, as confirmed by mass spectrometry (Wang et al., 2000). While this represents drop of more than 95% from wild-type DHFR expression (approx. 67 mg per litre of culture in rich media), this systems represents the first viable genetic code expansion in *E. coli*.

The *M. janaschii* system was further evolved for incorporation of photocrosslinking amino acid *p*-benzol-*L*-phenylalanine, using rounds of positive and negative selection (Chin et al., 2002a). Positive selection based on CAT assay identified the synthetase mutants that aminoacylated the $tRNA^{Tyr}_{CUA}$ based on survival in presence of the non-canonical amino acid. The positive selection was followed by a negative selection step, where mutants of TyrRS that aminoacylated the $tRNA^{Tyr}_{CUA}$ in the absence of the non-canonical amino acid (i.e. with one of the canonical amino acids) led to cell death via readthrough of the amber mutant of the barnase gene (Chin et al., 2002a). The positive/negative selection scheme used by Chin and colleagues has been subsequently used to evolve the *M. janaschii* TyrRS/$tRNA^{Tyr}_{CUA}$ pair for incorporation of a great variety of useful non-canonical amino acids, reviewed by Dumas and colleagues (Dumas et al., 2015). Unfortunately, the *M. janachii* TyrRS/$tRNA^{Tyr}_{CUA}$ system is not orthogonal with respect to the eukaryotic aaRSs and therefore is limited in its scope to genetic expansion in proakryotes (Fechter et al., 2001; Wang et al., 2000).

1.5.4 *Methanosarcinae* pyrrolysyl-tRNA synthetase/tRNA

While a range of aaRS/tRNA pairs have been tested for *in vivo* genetic ode expansion, the most versatile system has proven to be the *Methanosarcinae* pyrrolysyl-synthetase/tRNA pair. This pair occurs as an amber suppressor in a variety of methanogens and some bacteria (discussed in chapter 1.7.2) and is orthogonal in both prokaryotes and eukaryotes. Initial studies focused on analogue of pyrrolysine, which itself is difficult to synthesise (Polycarpo et al., 2006), and

showed that the PylRS/$tRNA_{CUA}$ from *Methanosarcina barkeri* pair is able to incorporate a wide variety of non-canonical amino acids (Li et al., 2009a; Li et al., 2009b; Polycarpo et al., 2006). Over the following years, many additional amino acids were incorporated using the wild-type aaRS/tRNA pair. The non-canonical amino acids encoded included lysine with Boc, azide and alkyne groups, where the latter two allow for site specific conjugation via copper catalysed [3+2] cycloaddition reactions (Nguyen et al., 2009a). Fekner and colleagues encoded a pyrrolysine analogue with terminal alkyne functionality that allowed similar site-specific labelling (Fekner et al., 2009), followed by two additional analogues reported later (Lee et al., 2013; Li et al., 2010).

The engineering efforts have been guided by structural characterisation of the PylRS/tRNA complex, which was provided by Kavran et al. in 2007. The authors used a pyrrolysine analogue *N*-ε-[(cylopentyloxy)carbonyl-*L*-lysine (cyc) to solve the structure of PylRS-AMP-cyc complex from *M. mazei* (Kavran et al., 2007). Taking advantage of the structural information and negative/positive selection protocol, the PylRS/tRNA pair has been evolved for incorporation of many different non-canonical amino acids. The first encoded amino acid using an engineered PylRS/tRNA was reported by Neumann and colleagues, who engineered PylRS mutant for incorporation of lysine bearing post-translational modification – the acetyl group (Neumann et al., 2008b). In their screen, Neumann et al. found a mutant of the *M. Barkeri* PylRS/tRNA with 6 mutations in the Pyl binding pocket that allowed for incorporation of *N*-ε-acetyllysine and, as shown subsequently, also its alkyl analogue 2-amin-8-oxononanoic acid (Huang et al., 2010). The incorporation of acetyllysine has been used extensively to study its biological role in histones and other proteins *in vitro* (Arbely et al., 2011; Lammers et al., 2010; Neumann et al., 2009) and *in vivo* (Di Cerbo et al., 2014; Tropberger et al., 2013).

The *Methanosarcina* PylRS/tRNA pair has been engineered for incorporation of wide variety of over 100 non-canonical amino acids (reviewed in (Dumas et al., 2015)). A particularly useful amino acids incorporated using the PylRS/tRNA pair were lysine derivatives with norbornene (Lang et al., 2012a), transcyclooctene and bicyclononyne (Lang et al., 2012b), which allow for copper-free, site-specific labelling in live cells. The system was also engineered for incorporation of an amino acid bearing the 2,2,5,5-tetramethyl-pyrrolin-1-oxyl spin label for paramagnetic distance measurements (Schmidt et al., 2014) and Virdee and colleagues

demonstrated incorporation of δ-thiol-L-lysine, which allowed to produce proteins with site-specific ubiquitination (Virdee et al., 2011).

In addition to amino acids for site-specific labelling and protein conjugation, a number of photocaged amino acids were incorporated using the PylRS/tRNA pair. The first photocaged amino acid reported was *o*-nitrobenzyloxycarbonyl-*N*-ε-L-lysine (Chen et al., 2009), followed by photocaged lysine used for protein location control (Gautier et al., 2010), light-activated transcription (Hemphill et al., 2013) and temporal dissection of MEK1 signalling (Gautier et al., 2011) in mammalian cells, and photocaged cysteine, used for photoactivation of TEV protease in mammalian cells (Nguyen et al., 2014).

Somewhat surprisingly, the PylRS/tRNA pair has been also successfully evolved for incorporation of phenylalanine analogues. The first demonstration was provided by Wang and colleagues, who successfully evolved the PylRS binding pocket to accommodate L-phenylalanune p-iodo-L-phenylalanine and p-bromo-L-phenylalanine (Wang et al., 2010). The pair was subsequently developed for incorporation of *O*-nitrobenzyl-*O*-tyrosine, a photocaged tyrosine derivative (Arbely et al., 2012), multiple O-substituted, meta-substituted and ortho-substituted tyrosine derivatives (Tharp et al., 2014; Wang et al., 2012a; Wang et al., 2012b), in addition to meta-alkoxy- and meta-acyl-phenylalanines (Tuley et al., 2014). The already broad range of non-canonical amino acids incorporated by the PylRS/tRNA pair was further extended to histidine derivatives by Xiao et al. (Xiao et al., 2014).

Two additional noteworthy *M. mazei* PylRS mutants were identified by Yanagisawa and colleagues through random mutagenesis library screening: Y384F and Y306A,Y384F (Yanagisawa et al., 2008). The Y384F mutant shows improved incorporation efficiency with BocK and AllocLys in *E. coli* (Yanagisawa et al., 2008) while the additional Y306A mutation allows for incorporation of a variety of non-canonical amino acids, including *N*-ε-benzyloxycarbonyl-*L*-lysine, *N*-ε-(*O*-azidobenzyloxycarbonyl-*L*-lysine (Yanagisawa et al., 2008), Se-alkylselenocysteines (Wang et al., 2012), N-e-[2-(furan-2-yl)ethoxy]carbonyl-lysine for light controlled protein-RNA crosslinking (Schmidt and Summerer, 2013), but also for bio-orthogonal handles for protein labelling, such as norbornene (Plass et al., 2011), transcyclooctenes (Plass et al., 2012) and cyclooctene derivatives (Borrmann et al., 2012).

Recently, a new aaRS/tRNA pair for genetic code expansion has been discovered in the species *Methanomethylophilus alvus* (Willis and Chin, 2018). Due to the increasing number of sequenced genomes, Willis and Chin have been able to identify number of PylRS/tRNA pair candidates based on a search of homology to the C-terminal catalytic domain of *M. mazei* PylRS and the absence of N-terminal domain of *Desulfobacterium hafniense*, which encodes PylRS as two distinct coding sequences that assemble to create functional synthetase in cell (Herring et al., 2007). This search yielded five candidate organisms for which both the PylRS gene and the corresponding $tRNA^{Pyl}$ gene were found. Of the five pairs, four were found to be orthogonal in *E. coli* and one was subsequently evolved to be orthogonal also with respect to the *M. mazei* PylRS/tRNA pair (Willis and Chin, 2018). This pair is discussed in detail in Chapter 4.

1.5.5 Post-translational modifications and phosphoseryl-tRNA Synthetase/tRNA pair

One of the major efforts within genetic code expansion is the efficient genetic encoding of amino acids bearing post-translational modifications (Chin, 2017). Access to site-specifically post-translationally modified proteins provides researchers a unique tool to understand the role of post-translational modifications, many of which play a prominent role in cell physiology and pathology (reviewed in section 1.3). Generally, post-translational modifications are installed enzymatically and are tightly regulated within cells. While it may be possible to post-translationally modify proteins *in vitro* by a known enzyme, the effects are often not site-specific and are difficult to generalise to all proteins. An alternative to phosphorylation is the so-called phosphomimetic mutations of the given residue to aspartate or glutamate; however, these amino acids are distinct and often fail to recapitulate the given phenotype (Roberts-Galbraith et al., 2010).

The first amino acid bearing a post-translational modification genetically encoded via the *M. barkeri* PylRS-tRNA was *N*-ε-acetyl-lysine (Neumann et al., 2008b). Subsequently, many post-translational modifications have been encoded via the PylRS/tRNA pair, including lysine methylation (Groff et al., 2010; Nguyen et al., 2010; Nguyen et al., 2009a; Wang et al., 2010) and lysine ubiquitination (Li et al., 2009b; Virdee et al., 2011; Virdee et al., 2010), as well as

M. janaschii TyrRS/tRNA pair, which has been successfully used for incorporation of Phosphotyrosine and it's analogues (Fan et al., 2016; Hoppmann et al., 2017; Luo et al., 2017; Xie et al., 2007).

Recently, the phosphoseryl-tRNA synthetase from *Methanococcus mariplaudis* in combination with $tRNA^{Cys}$ *Methanocaldococcus janaschii* from was adapted for the direct genetic encoding of phosphoserine (Park et al., 2011). In certain methanogens the $tRNA^{Cys}$ is first aminoacylated by phosphoserine, which is subsequently converted to cysteine via action of the SepCysS enzyme (Sauerwald et al., 2005). Park et al. first converted the *Mj*$tRNA^{Cys}$ to an amber suppressor pair, via mutation of the anticodon and an additional mutation to improve aminoacylation (Hohn et al., 2006). In addition, the authors have evolved the bacterial elongation factor EF-Tu to improve the shuttling of the charged Sep-$tRNA^{Cys}$ to the ribosome. Combination of the two factors allowed for production of low levels of proteins with site-specific phosphoserine (Park et al., 2011).

Because SepRS recognises recognises the $tRNA^{Cys}$ anticodon (Fukunaga and Yokoyama, 2007), this interaction is compromised by converting the anticodon into amber suppressor, resulting in inefficient system for incorporation of phosphoserine. Rogerson and colleagues managed to significantly improve the system through multiple rounds of directed evolution, optimising first the tRNA antiocodon stem loop and subsequently the SepRS/tRNA interface, identifying a pair that enabled yields of several milligrams of protein per litre of culture (Rogerson et al., 2015).

In 2016, Yang and colleagues have used the SepRS/tRNA pair to incorporate phosphoserine into cells and subsequently removed the phosphate, converting the residue to dehydroalanine. Using the zinc and copper mediated addition of alkyl iodides, the authors were able to install several post-trasnlational modifications including various methyllysines, formyllysine and acetyllysine. However, this approach suffers from limited selectivity (80%) and yield (50%) (Yang et al., 2016).

Zhang and colleagues have evolved the SepRS system for incorporation of phosphothreonine (Zhang et al., 2017). In their work, Zhang and colleagues first genetically engineered *E. coli* to produce high level of intracellular phosphothreonine by importing the *pduX* gene from

Salmonella eneterica, which codes for a kinase that converts L-threonine into phosphothreonine (Fan and Bobik, 2008; Fan et al., 2009). Subsequently, the $tRNA^{v1.0}_{CUA}$ ($tRNA^{Cys}_{CUA}$ reported by (Rogerson et al., 2015)) was evolved to improve orthogonality with respect to endogenous synthetases, identifying a mutant designated $tRNA^{v2.0}_{CUA}$ with 4,000-fold lower level of amber readthrough in the absence of SepRS (Zhang et al., 2017). The $tRNA^{v2.0}_{CUA}$ was used in a new directed evolution protocol to change the substrate specificity for the SepRS reported by Rogerson et al. The protocol uses parallel positive selection in presence and absence of non-canonical amino acid (i.e. phosphothreonine) and statistical comparison of surviving clones, analysed using deep sequencing (Zhang et al., 2017). The process led into identification of 16 clones that are selectively enriched in presence of phosphothreonine, one of which led to quantitative incorporation of phosphothreonine, designated pThrRS/$tRNA^{v2.0}_{CUA}$. This pair allowed Zhang et al. to express recombinant proteins for both structural and biochemical studies (Zhang et al., 2017).

1.6 Limitations of *in vivo* genetic code expansion

While great effort is directed toward making the engineered aaRS/tRNA pairs orthogonal with respect to the host endogenous aaRS/tRNAs, biological orthogonality cannot be absolute. The overall endogenous translation error rate is estimated to be approximately10^{-4} (Loftfield and Vanderjagt, 1972, Kramer and Farabaugh, 2007) and misaminoacylation of endogenous tRNAs by the endogenous synthetases is known to contribute to this error rate (Söll, 1990)(Söll, 1990). Similarly, the UAG codon is misread by endogenous tRNAs at a measureable level (Roy et al., 2015). The reported error rates of new orthogonal pairs are less than 10^{-3} for optimised and widely used aaRS/tRNA pairs (measured either *in vitro* or by determining misincoporation at sense codons *in vivo*) (Liu and Schultz, 2010; Neumann et al., 2008b; Zhang et al., 2017).

While amber codon suppression solves the problem with no available sense codons for incorporation of non-canonical amino acids, this solution leads to some loss of fitness caused by readthrough of the endogenous stop codons (Aerni et al., 2015). The Church group has attempted to address this issue in *E. coli* by removal of all of the 314 amber stop codons from its genome. The approach combined multiplex automated genome engineering (MAGE), which

relies on introduction of oligo-mediated allelic replacement through λ-red mediated recombination (Wang et al., 2009) with conjugation based gene transfer. First, 32 *E. coli* strains were engineered via MAGE removing the TAG codons from different sections of the genomes. These strains were subsequently combined together using an conjugation based hierarchical assembly to produce first 8 strains, each with 1/8 of their genome recoded (Isaacs et al., 2011) and subsequently into a single strain (Lajoie et al., 2013). The strain allowed for deletion of the RF1, however, additional 355 mutations were acquired in the assembly process. Ostrov and colleagues used this strategy to test a 57-codon recoding scheme, removing the amber stop codon and additional synonymous sense codons, in 55 individual segments (Ostrov et al., 2016). An alternative strategy was developed by Wang and colleagues, who used Cas9 in combination with negative and positive selection markers to recode large sections of the genome (~ 100 kb) at a time (Wang et al., 2016). In their report, Wang and colleague demonstrate a feasible recoding schemes, removing subsets of sense codons from a subsection of the *E. coli* genome (Wang et al., 2016).

1.6.1 Prokaryotic release factor engineering

The use of amber codon as a codon for incorporation of non-canonical amino acid has intrinsic limitations. As discussed in previous chapter, readthrough of endogenous amber codons may lead to decreased cellular fitness. However, the incorporation of non-canonical amino acid via amber suppressor $tRNA_{CUA}$ also leads to a competition in between the $tRNA_{CUA}$ and RF1, which is recruited to the stalled ribosome whenever the target UAG codon is encountered (discussed in section 1.2.1), resulting in a truncated protein released from the ribosome. This limitation has led to efforts to either increase the intracellular $tRNA_{CUA}$ concentration or engineering of the bacterial RF1.

Due to the overlapping specificity of the bacterial release factors (RF1 terminates translation at UAA and UAG codons, while RF2 terminates translation at UAA and UGA), only RF1 needs to be engineered. While an essential gene (Rydén and Isaksson, 1984), several groups managed to delete the *prfA* gene while compensating for the activity of RF1. Mukai and colleagues have cloned 7 essential genes (*coaD*, *had*, *hemA*, *mreC*, *murF*, *lolA and lpxK*) that end with TAG codon onto an artificial bacterial chromosome (BAC) and substituted the stop

codons for UAA and introduced an amber suppressor $tRNA^{Gln}_{CUA}$, which in combination with the BAC sufficed to complement for the deletion of the *prfA* gene (Mukai et al., 2010a). An alternative strategy was devised by Johnson and colleagues who have introduced two mutations into the *prfB* gene which codes for RF2. The resulting release factor was sufficient for termination at all stop codons, permitting the deletion of *prfA* gene (Johnson et al., 2011).

1.6.2 Quadruplet codon

An alternative to amber codon suppression is the use of quadruplet codons. Quadruplet codons have been first examined in the context of genetic code expansion used by Hohsaka and colleagues, who used tRNAs with extended anticodon loops (for incorporation of nitrophenylalanine through *in vitro* translation, proving that ribosomal decoding of quadruplet codon is possible (Hohsaka et al., 1996).

The first *in vivo* experiments demonstrating quadruplet codon decoding were done by Moore et al., who have shown that $tRNA^{Leu}$ from *E. coli* could be engineered to recognise UAGA by inserting an extra nucleotide U at position 33.5 of its anticodon loop (Moore et al., 2000). This effort was followed by Magliery and colleagues, who have engineered several additional variants of tRNA with quadruplet anticodon using the *E. coli* $tRNA^{Ser}$ as a base for a library screen (Magliery et al., 2001).

1.6.3 Ribosome engineering

The development of orthogonal ribosome has enabled additional significant advances in genetic code expansion. The difficulty of engineering an essential translational machinery is evident – any modification that interferes with the endogenous function very likely has a negative phenotype. Rackham and Chin elegantly circumvented this problem by developing a network of orthogonal ribosome and mRNA pairs based on altered Shine-Dalgarno sequence. This allowed a specific subset of orthogonal mRNA and orthogonal small ribosomal subunit to create a functional pair inside the cell without interaction with any of the endogenous mRNAs and ribosomes (Rackham and Chin, 2005). The orthogonal ribosome was subsequently evolved by Wang and colleagues for increased amber codon suppression through decrease in the

interaction between the orthogonal small ribosomal subunit and the RF1 (Wang et al., 2007a). Using the orthogonal ribosome first reported by Wang et al., Neumann and colleagues evolved the small ribosomal subunit for efficient decoding of quadruplet codons (Neumann et al., 2010)

1.7 Expansion of the genetic code in eukaryotic cells

Parallel to the efforts in genetic code expansion in *E. coli*, researchers have focused on expanding the genetic code of eukaryotic organisms (Chin, 2014). Similarly to expanding the prokaryotic code, the ability to incorporate non-canonical amino acids would allow scientists to probe, control and image protein function *in vivo* in addition to providing a route for new therapeutic approaches (Chin, 2017).

Saccharomyces cerevisiae was the first organism with successfully expanded genetic code (Chin et al., 2003). Chin and colleagues used the *E. coli* tyrosyl-tRNA synthetase/$tRNA_{CUA}$ pair, with the anticodon mutated to recognise the amber codon. The *E. coli* tyrosyl-tRNA synthetase is known to be orthogonal in yeast, i.e. the TyrRS aminoacylates the *E. coli* $tRNA^{Tyr}_{CUA}$ but not any of the endogenous yeast tRNAs (Edwards et al., 1990; Edwards et al., 1991) and similarly, the *E. coli* $tRNA^{Tyr}_{CUA}$ is not aminoacylated by endogenous yeast synthetases (Trezeguet et al., 1991). In addition, the TyrRS does not have an editing domain (Fersht et al., 1980; Jakubowski and Goldman, 1992), making it an attractive candidate for altering the substrate specificity. The authors used the positive and negative selection cycle in presence and absence of the non-canonical amino acid, respectively, to select active mutants from library of TyrRS with 5 randomised residues. Positive selection was based on amber mutants of transcriptional activator GAL4 (that activated transcription of HIS3, *URA3* and *lacZ* genes) and carried out in absence of uracil or in presence 3-aminotriazole and lacking histidine. The negative selection was subsequently carried out in presence of 5-fluorootic acid (Chin et al., 2003). The authors were able to encode five different non-canonical amino acids in yeast, including *p*-acetyl-L-phenylalanine, *p*-benzoyl-L-phenylalanine, *p*-azido-L-phenylalanine, O-methyl-L-tyrosine and *p*-iodo-L-phenylalanine (Chin et al., 2003). This system has been subsequently used in photocrosslinking studies of the RNA polymerase II (Chen et al., 2007).

The first addition of non-canonical amino acid in mammalian cells was reported by Sakamoto et al., who have discovered that *Bacillus stearothermophilus* suppressor $tRNA^{Tyr}_{CUA}$ (*Bst*$tRNA^{Tyr}_{CUA}$) contains an internal promoter sequence, allowing for its expression in mammalian cells. The authors then combined the $tRNA^{Tyr}_{CUA}$ with *E. coli* TyrRS (*Ec*TyrRS) mutant (expressed from tetracycline-regulated promoter), which allowed the incorporation of 3-iodo-L-tyrosine into Ras protein (Sakamoto et al., 2002). In second example, *Bacillus stearothermophilus* tryptophanyl-tRNA (*Bst*$tRNA^{Trp}$) was converted to opal supressor and used in combination with its cognate tryptophanyl-tRNA synthetase mutated to accept 5-hydroxy-L-tryptophan, a fluorescent probe that undergoes chemical oxidation *in situ*, allowing for protein cross-linking in mammalian cells (Zhang et al., 2004).

In addition to being orthogonal in yeast, the *Bst*$tRNA^{Tyr}_{CUA}$ with the *Ec*TyrRS pair is also orthogonal in mammalian cells due to a high degree of conservation of the aaRS/tRNA pairs among eukaryotes. This was first demonstrated by Liu et al. who have combined the *Bst*$tRNA^{Tyr}_{CUA}$ with the *Ec*TyrRS mutants discovered by Chin and colleagues (Chin et al., 2003) and optimised their expression in mammalian cells to incorporate six different non-canonical amino acids in mammalian cells (Liu et al., 2007). Wang and colleagues have further demonstrated that the expression of exogenous suppressor $tRNA_{CUA}$ can be driven by type-3 pol III promoter, using the *Ec*TyrRS/*Ec*$tRNA^{Tyr}_{CUA}$ to incorporate different non-canonical amino acids into K+ channel Kv1.4 in mammalian HEK cells and neurons using transient transfection (Wang et al., 2007b).

Analogous to genetic code expansion in *E. coli*, the imported aaRS/tRNA pair has to be orthogonal in the eukaryotic host. While the *E. coli* TyrRS/tRNA is orthogonal in eukaryotes, it has to be evolved directly in the eukaryotic organism due to lack of orthogonality in *E. coli* where library-screening approaches are easier to execute. This situation has limited the use of *E. coli* aaRS/tRNAs for genetic code expansion in eukaryotes. Conversely, the *M. janaschii* TyrRS/$tRNA^{Tyr}_{CUA}$ pair has proven to be useful for incorporation of diverse set of non-canonical amino acids in *E. coli* where it can be also evolved, however, the pair is not orthogonal in eukaryotic cells.

This situation has led to the adaptation of the PylRS/tRNA pair, which is orthogonal in both *E. coli* and all eukaryotic hosts tested to date. This characteristic allows researchers to evolve the pair PylRS/tRNA for a specific substrate in *E. coli* and then transfer the pair to the eukaryotic host. The first use of the PylRS/tRNA pair in yeast was reported by Hancock et al., who have incorporated a variety of non-canonical amino acids, including amino acid with an alkyne group, a photocaged amino acid and photo-cross-linking amino acid (Hancock et al., 2010). Testing out multiple possible promoter sequences, the group found that the A- and B-box promoters in $tDNA^{Arg}_{UCU}$ gene drive sufficient expression of the $tRNA^{Pyl}_{CUA}$ gene when it is substituted instead of the $tDNA^{Arg}_{UCU}$ gene. Using this construct, the authors were able to successfully express in yeast cells multiple PylRS/tRNA pairs initially evolved in *E. coli* (Hancock et al., 2010). The PylRS/tRNA pair has been since then used for incorporation of variety of substrates including non-canonical amino acids for protein labelling (Lang et al., 2012a; Nguyen et al., 2011, Borrmann et al., 2012; Lang et al., 2012b; Plass et al., 2012; Ramil et al., 2017), photocrosslinking (Chou et al., 2011; Zhang et al., 2011), and photoactivation (Arbely et al., 2012; Gautier et al., 2011) in mammalian cells.

Recently, Italia and colleagues reported a new strategy for evolution of *E. coli* aaRS/tRNA pairs directly in the *E. coli* cell (Italia et al., 2017). The authors took advantage of previous efforts to replace an endogenous *E. coli* aaRS/tRNA pair with an evolutionarily distant pair (Iraha et al., 2010) and thereby liberating the given endogenous pair for directed evolution. Selecting the tryptophanyl-tRNA synthetase as a useful potential target, the authors first replaced the endogenous *Ec*TrpRS/tRNA pair with *S. cerevisiae* (*Sc*) TrpRS/tRNA, which is orthogonal in *E. coli* (Hughes and Ellington, 2010), at the endogenous locus and then converted the *Ec*$tRNA^{Trp}$ into amber suppressor. In the following step, the *Ec*TrpRS/$tRNA_{CUA}$ pair was evolved using positive and negative selection for incorporation of 5-hydroxytryptophan. In addition, the isolated *Ec*TrpRS mutants were tested for incorporation of other 5-substituted tryptophan derivatives (Italia et al., 2017).

1.8 Expansion of the genetic code in animals

Function and development of complex, multi-cellular biological structures, such as the brain, the nervous system and the immune system can be only studied in animals. This has motivated researchers to further extend the genetic code expansion into animals, where the powerful technology can be used to answer critical biological questions (Chin, 2014).

The first demonstration of genetic code expansion in an animal was achieved in the nemoatode *Caenorhabditis elegans* (*C. elegans*) (Greiss and Chin, 2011). *C. elegans* is a model organism with several advantages which make it an ideal candidate for genetic code expansion. Firstly, the entire lineage of every cell in the development of this nematode has been mapped (Kimble and Hirsh, 1979; Sulston and Horvitz, 1977). In addition, *C. elegans* is transparent throughout its entire life cycle, allowing for very efficient use of photosensitive non-canonical amino acids. In addition, amber suppressor mutations have been isolated and characterised in *C. elegans* (Hodgkin, 1985; Kondo et al., 1988; Kondo et al., 1990) suggesting that amber suppression is not detrimental to survival of the organism.

Greiss et al. achieved incorporation of two non-canonical amino acids: *N*6-(*tert*-butyloxycarbonyl)-L-lysine and N6-[(2-propynyloxy)carbonyl]-L-lysine; into proteins in *C. elegans*. The authors used the PylRS/tRNA pair from *M. mazei*, which is orthogonal in both *E. coli*, yeast and mammalian cells. The PylRS gene was expressed using a ubiquitous rps-0 promoter while the tRNA was expressed using extragenic RNA polymerase III promoter (Greiss and Chin, 2011).

Since genetic code expansion via suppression of the amber codon requires insertion of the TAG amber stop codon into the middle of the coding sequence, i.e. far from the 3' mRNA end, the mRNA message of the target protein is subject to nonsense mediated decay pathway (NMD). NMD is a quality-control process that protects the organism from mRNAs with a premature stop codon by their degradation (Chang et al., 2007; Longman et al., 2007). To investigate whether NMD was a major factor, Greiss et al. created worms bearing mCherry-TAG-GFP reporter and crossed them with worms deficient in *smg-2(e2008)* worms that are deficient in NMD (Hodgkin et al., 1989; Page et al., 1999). The Δsmg-2 worms showed significantly higher

expression of the full-length product in the presence of the non-canonical amino acid, as measured by GFP fluorescence, suggesting that NMD can be an important factor in genetic code expansion in eukaryotes and animals (Greiss and Chin, 2011). A similar approach showed subsequently that *E. coli* tyrosyl- and leucyl-RS/tRNA pairs can also be converted to amber suppressors and used for genetic code expansion in *C. elegans* (Parrish et al., 2012).

The fruit fly *Drosophila melanogaster* was the second model organism with expanded genetic code (Bianco et al., 2012). Mukai and collegues have demonstrated in 2010 that *Ec*TyrRS/tRNA is orthogonal in *Drosophila* melanogaster Schneider 2 cells in culture and can be used to incorporate 4-azido-L-phenylalanine into interleukin-8 using amber suppression (Mukai et al., 2010b). Similarly to *C. elegans*, amber suppressors in the fly have been reported (Garza et al., 1990; Laski et al., 1989; Washburn and O'Tousa, 1992) and are therefore compatible with survival of all developmental stages. The authors expressed the *M. mazei* PylRS using the GAL4 upstream activating sequence and $tRNA^{Pyl}$ by extragenic Pol III promoter (Bianco et al., 2012), an approach that allows for crosses with a variety of existing GAL4 lines that should enable expression in variety of tissues. Bianco et al. demonstrated that this approach can be used for incorporation of several non-canonical amino acids in tissue-specific manner, including amino acids with norbornene and alkyne groups for site-specific protein labelling (Bianco et al., 2012). The genetic code expansion in *D. melanogaster* via the PylRS/tRNA pair was successfully used by Elliott and co-workers to label and image proteins in specific tissues and developmental stages (Elliott et al., 2014).

The remainder of this thesis discusses several key improvements in genetic code expansion in mammalian cells. Chapter 2 discusses the first genetically encoded phosphorylation in mammalian cells, Chapter 3 presents technological progress in protein labelling technology and its application to novel class of microproteins and Chapter 4 introduces a new, orthogonal pair in mammalian cells, paving the way forward for simultaneous incorporation of multiple distinct non-canonical amino acids.

Chapter 2 Genetically Encoded Protein Phosphorylation in Mammalian Cells

2.1 Introduction

This chapter describes the application of the recently engineered phosphoseryl-tRNA synthetase/tRNA pair ($SepRS^{v1.0}$/$tRNA^{v1.0}_{CUA}$) in mammalian cells for co-translational incorporation of phosphoserine and its non-hydrolysable phosphonate analogue.

As introduced in section 1.4, post-translational modifications extend the chemical space of proteins by chemically modifying the side chains of individual amino acids or the protein termini. Post-translational modifications have been implicated in of almost all biological functions in eukaryotes (Walsh et al., 2005) and are increasingly more appreciated in bacteria (Grangeasse et al., 2015). The approaches to study the role of post-translational modifications are limited due to the fact that many of the enzymes responsible for particular modifications are unknown or may not modify the target protein site-specifically or completely.

Genetic code expansion is a unique tool for investigation of post-translational modifications, providing researchers access to site-specifically modified proteins. To date, researchers have successfully encoded number of amino acids containing post-translational modification, including methylation (Ai et al., 2010; Nguyen et al., 2009a), acetylation (Neumann et al., 2008b), crotonylation (Kim et al., 2012), sulphation (Liu and Schultz, 2006), nitration (Neumann et al., 2008a) and phosphorylation (Fan et al., 2016; Park et al., 2011; Rogerson et al., 2015; Zhang et al., 2017).

The incorporation of post-translationally modified tyrosine was achieved primarily through directed evolution of the *Methanococcus janaschii* (*Mj*) tyrosyl-tRNA synthetase (TyrRS)/tRNA pair. In 2006, Liu and Schultz first reported the genetic encoding of

sulfotyrosine using amber suppressor *Mj*tRNA$^{Tyr}_{CUA}$ (Liu and Schultz, 2006). The Schultz group subsequently reported incorporation of phosphotyrosine analogue p-carboxymethyl-L-phenylalanine (Xie et al., 2007), and Neumann and colleagues successfully evolved the to incorporate 3-nitrotyrosine into manganese superoxide dismutase (Neumann et al., 2008a).

Lysine bearing a variety of post-translational modifications was encoded using the pyrrolysine-tRNA synthetase (PylRS)/tRNA pairs from *Methanosarcina mazei* (*Mm*) and *Methanosarcina barkeri* (*Mb*). The first encoded lysine bearing post-translational modification was Nε-acetyllysine, reported in 2008 by Neumann and colleagues (Neumann et al., 2008b). This effort was followed by Nguyen et al., who used the *Mb*PylRS/tRNA$_{CUA}$ pair to encode Nε-methyl-L-lysine. To achieve selectivity over the endogenous lysine, which differs by only the absence of single methyl group, Nguyen et al. first encoded the Nε-*tert*-butyl-oxycarbonyl-L-lysine and removed the *tert*-butyl-oxycarbonyl group in presence of 2% trifluoroacetic acid (Nguyen et al., 2009a). An analogous approach was demonstrated by Ai et al., who have genetically encoded Nε-allyloxycarbonyl-Nε-methyl-L-lysine, which can be converted to methyl-L-lysine using ruthenium catalyst (Ai et al., 2010) and Wang and colleagues, who have encoded a photocaged Nε-methyl-L-lysine variant (Wang et al., 2010). A strategy to produce proteins bearing ε-N, N-dimethyl-L-lysine was subsequently reported by Nguyen and colleagues, who have used the incpororation of Nε-*tert*-butyl-oxycarbonyl-L-lysine in combination with protection of the all amine groups via N-(benzyloxycarbonyl)succinimide reductive methylation, and subsequent amine deprotection (Nguyen et al., 2010). Since this approach relies on denaturing conditions of the succinimide protection, Wang and colleagues have developed an alternative approach, encoding Nε-(4-azidobenzoxycarbonyl)-δ,ε-dehydro-L-lysine, which is can be converted to dimethyllysine in two biocompatible reactions (Wang et al., 2017).

The approaches using co-translational incorporation of amino acids bearing post-translational modifications are complimented by incorporation of non-canonical amino acids which can be used as chemical handles that allow for subsequent derviatization (reviewed in (Krall et al., 2016; Lang and Chin, 2014)). Notably, post-translational modifications including, glycosylation, ubiquitination or phosphorylation were achieved using this approach. The first example of glycosylation was reported by Liu and colleagues who have incorporated *p*-acetyl-

L-phenylalanine that was subsequently conjugated with β-linked analogue of N-acetylglucosamine (Liu et al., 2003). The approach has been extended by incorporation of biochemical handles such as alkene- and alkyne-lysine for copper mediated azide-alkyne cycloaddition, for derivatization of lysine with multiple sugar modifications (Kaya et al., 2009). Similar approach using Staudinger-phosphite reaction of genetically encoded azido-phenylalanine, resulted in an non-canonical amino acid mimicking phosphotyrosine (Serwa et al., 2009).

Site-specific protein ubiquitination was achieved by incorporation of Nε-cysteinyl-L-lysine combined with native chemical ligation (Li et al., 2009b), however, this approach results in a non-native linkage, which cannot be recognised by deubiquitinases, limiting the use of this technique. An alternative approach termed genetically encoded orthogonal protection and activated ligation (GOPAL) using Nε-Boc-L-lysine combined with protection of the remaining lysine residues with N-(benzyloxycarbonyloxy)-succinimide and conjugation to ubiquitin thioester (Virdee et al., 2010), which, unlike the previously reported strategies, results in native ubiquitin linkage but does rely on protein refolding after denaturing conditions of the N-(benzyloxycarbonyloxy)-succinimide protection. A different traceless approach is based on incorporation of δ-thiol-Nε-(*p*-nitrocarbobenzyloxyl)-L-lysine, which is post-translationally deprotected to δ-thiol-L-lysine and the resulting sidechain can react with ubiquitin thioester (Virdee et al., 2011).

Recently, significant efforts have been directed towards genetic encoding of phosphorylated amino acids (Chin, 2017). Incorporation of phosphotyrosine has been reported by several groups. Fan et al. used the *Mj*TyrRS/tRNA pair and an evolved elongation factor Tu in an *E. coli* strain with several deleted phosphatases to incorporate phosphotyrosine into GFP reporter, but with only limited yield and mixture of amino acids incorporated at the target position (Fan et al., 2016). Using an alternative strategy, Hoppmann and colleagues developed a system for incorporation of phosphoramidate, which is converted to phosphotyrosine at low, denaturing pH (Hoppmann et al., 2017). Simultaneously, Luo et al. have reported a propeptide strategy for increase of cellular uptake of phosphotyrosine and 4-phosphomethyl-L-phenylalanine, a non-hydrolyzable phosphotyrosine analogue, which allowed them to identify a *Mj*TyrRS/tRNA pair with sufficient efficiency for phosphotyrosine incorporation (Luo et al., 2017).

In parallel to advances in genetic encoding of post-translational modifications, several groups pioneered the expansion of the genetic code of mammalian cells and animals (reviewed in sections 1.7 and 1.8). The combination of these two domains would provide researchers the ability to install post-translationally modified amino acids *in vivo*, allowing them to study their effect within the cellular context.

One of the first reports of encoding Nε-acetyllysine was provided by Yokoyama groupp in 2008 (Mukai et al., 2008), followed by Groff et al., who reported encoding of a photocaged version of Nε-methyl-L-lysine (Groff et al., 2010). In a more recent example, Elsasser et al. combined a new mammalian expression system (Schmied et al., 2014) with transposase mediated stable cell line selection to incorporate Nε-acetyl-L-lysine into histones in mouse embryonic stem cells and were able to observe the effects in gene expression (Elsässer et al., 2016).

The most prevalent post-translational modification – among the more than 300 known to date (Walsh, 2006) – is protein phosphorylation, with over 260,000 experimentally verified phosphorylation sites (www.phosphosite.org). The human genome contains 518 known kinases, amounting to approximately 1.7% of all human genes (Manning et al., 2002). Previous efforts to genetically encode phosphorylated amino acids, summarised in section 1.6.5, have resulted in an optimised system that enables high expression of proteins bearing phosphoserine in *E. coli* (Rogerson et al., 2015). This system allows researchers to purify quantitatively and site-specifically phosphorylated proteins for both *in vitro* and structural studies but does not allow the study of phosphorylation consequences in their native context. Following this development, Zhang et al. have reported a new, evolved version of the SepRS/tRNA$^{Sep}_{CUA}$ pair that, together with metabolic engineering of the bacterial cell, enables the incorporation of phosphothreonine in proteins (Zhang et al., 2017). These advances have yielded significant insights into the role of protein phosphorylation, including the role of ubiquitin phosphorylation (Huguenin-Dezot et al., 2016) and protein folding (Mukherjee et al., 2018).

This chapter describes the first genetic encoding of phosphorylation in mammalian cells, allowing researchers to study protein phosphorylation in a site-specific manner directly in cells. To achieve this, we completed several steps described in the following sections:

1. We transferred the SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ system for site-specific installation of phosphorylated amino acid into expression system for mammalian cells;
2. We engineered the eukaryotic elongation factor 1 α (EF-1α) to increase the amber suppression efficiency;
3. We combined the resulting system with the engineered eukaryotic release factor 1 (eRF1) for increased amber suppression;
4. We increased the intracellular concentration of phosphoserine by knocking out phosphoserine phosphatase.

All of the above steps lead to an order of magnitude increase in encoding efficiency over non-optimised system. Because post-translational modifications are dynamic and tightly regulated *in vivo*, the genetically encoded amino acids may be subject to post-translational dephosphorylation by one of the endogenous protein phosphatases. To address this issue, we

5. Engineered the cell metabolic pathway by knocking out phosphoserine aminotransferase to decrease the cellular concentration of phosphoserine;
6. Demonstrate that the system incorporates a non-hydrolysable phosphonate analogue of phosphoserine and, lastly;
7. Demonstrate that incorporation of phosphonate analogue into a known activation site in protein kinase leads to its activation.

Together this chapter establishes genetic code expansion as a useful tool for studying protein phosphorylation in mammalian cells in a site-specific manner. In addition, the work establishes the SepRS/tRNA^{Sep} pair as an independent orthogonal pair for genetic code expansion in eukaryotes.

The following section has been published as:

Beránek, V., Reinkemeier, C.D., Zhang, M.S., Liang, A.D., Kym, G. and Chin, J.W., 2018. Genetically Encoded Protein Phosphorylation in Mammalian Cells. *Cell Chemical Biology*, *25*, 1–8.

The following list specifies my contributions to the manuscript as well as the contributions of the other co-authors:

J.W.C. defined the direction of research. V.B. and C.D.R. designed and performed the experiments. M.S.Z. performed the *in vitro* aminoacylation experiments. A.D.L. performed the measurements of intracellular amino acid concentrations (all LC/MS data). G.K. designed the initial plasmids. V.B. and J.W.C. wrote the paper with input from all authors. The individual contributions are detailed in figure legends.

Supplementary figures were inserted into the text where first mentioned and the numbering of figures has been modified from the publication.

2.2 Genetically encoded protein phosphorylation in mammalian cells

2.2.1 Introduction

Protein phosphorylation is a key post-translational modification that expands the complexity of protein function, and regulates diverse biological processes in eukaryotic systems (Manning et al., 2002). Thousands of phosphorylation sites have been discovered (Olsen et al., 2006), but the consequences of protein phosphorylation can be hard to determine, as manipulation of the kinases that install phosphorylation commonly have pleiotropic effects, and so-called phosphomimetic mutations to aspartic acid and glutamic acid are chemically distinct (Mandell et al., 2007) and often fail to recapitulate the molecular and phenotypic consequences of phosphorylation (Roberts-Galbraith et al., 2010).

Genetic code expansion, using orthogonal aminoacyl-tRNA synthetase/$tRNA_{CUA}$ pairs, has enabled the site-specific incorporation of diverse non canonical amino acids (ncAAs) into proteins in cells and animals (Chin, 2014). One key application of this approach is in understanding the role of post-translational modifications in regulating the biological functions of proteins. Directly genetically encoding ncAAs, corresponding to post-translationally modified versions of natural amino acids, into proteins has enabled the synthesis of recombinant proteins bearing defined post translational modifications, and led to numerous insights into how these modifications regulate protein structure and function (Davies and Chin, 2012, Chin, 2017).

Recently, a phosphoseryl-tRNA synthetase (SepRS)/$tRNA_{GCA}$ pair that directs the first step in a tRNA dependent cysteine biosynthesis pathway in certain methanogens (Sauerwald et al., 2005) has been co-opted for the site-specific incorporation of phosphoserine (**1**) into recombinant proteins in *E. coli*. (Park et al., 2011; Rogerson et al., 2015). Since SepRS recognizes the anticodon of $tRNA_{GCA}$, the SepRS/$tRNA_{CUA}$ pair - in which the GCA anticodon was simply substituted by CUA - was a very inefficient amber suppressor. In recent work, we demonstrated that phosphoserine could be efficiently incorporated into proteins in *E. coli* using an evolved SepRS/$tRNA_{CUA}$ pair (Rogerson et al., 2015). This pair, in which SepRS and the anticodon stem and anticodon loop of $tRNA_{CUA}$ were evolved to function efficiently, referred

to herein as the SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ pair, has been used to produce a number of site-specifically phosphorylated proteins for structural and functional studies (Rogerson et al., 2015; Huguenin-Dezot et al., 2016, Burgess et al., 2018, Dickson et al., 2018). We also demonstrated that by manipulating phosphoserine biosynthesis in *E. coli* a non-hydrolyzable, phosphonate analog of phosphoserine (**2**) could be incorporated into proteins using the SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ pair (Rogerson et al., 2015), and that the pair can be further evolved to incorporate phosphothreonine in *E. coli* (Zhang et al., 2017).

The ability to encode phosphoserine, and its non-hydrolyzable analogs, into defined sites in proteins in mammalian cells would facilitate an understanding of the molecular and cellular consequences of this modification. Unlike approaches that manipulate kinases and phosphatases, that have many targets in the cell, orthogonal routes to installing site-specific phosphorylation may directly address the consequences of modifying a particular site on a particular protein.

Orthogonal routes to installing other post-translational modifications have begun to emerge. We recently explored the genetic encoding of acetyl-lysine into chromatin (Elsässer et al., 2016), and complementary work explored directing protein ubiquitination into chromatin via protein trans splicing (David et al., 2015); both these approaches modify only a small fraction of histones but have led to important conclusions about the consequences of these modifications for gene expression and cellular interactions.

Protein phosphorylation is commonly activating and phosphoproteomics studies reveal that proteins are naturally regulated by sub-stoichiometric phosphorylation, with occupancies of less than 30% reported on the majority of phosphorylation sites (Olsen et al., 2010). Therefore, routes to the sub-stoichiometric installation of phosphorylated amino acids into proteins in mammalian cells will be useful.

Here we demonstrate that the SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ pair is orthogonal in mammalian cells and can be used, in combination with other engineered translational components, to site-specifically and co-translationally incorporate phosphoserine and its non-hydrolyzable phosphonate analog into proteins expressed in mammalian cells. Moreover, we demonstrate the synthetic activation of a protein kinase in mammalian cells by encoding a phosphonate analog at a site that is normally phosphorylated by an upstream kinase.

2.2.2 Results

2.2.2.1 Expressing SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ in human cells

We first demonstrated that the SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ pair can be expressed in mammalian cells. We were able to detect FLAG tagged SepRS$^{v1.0}$, following transfection of HEK293 cells with the corresponding gene, by western blot (**Figure 2.1a**). Similarly, upon expressing four repeats of the gene encoding tRNA$^{v1.0}_{CUA}$ from extragenic U6 promoters we detected expression of tRNA$^{v1.0}_{CUA}$ by northern blot (**Figure 2.1a**). These experiments demonstrate that SepRS and tRNA$^{v1.0}_{CUA}$ can be expressed in human cells.

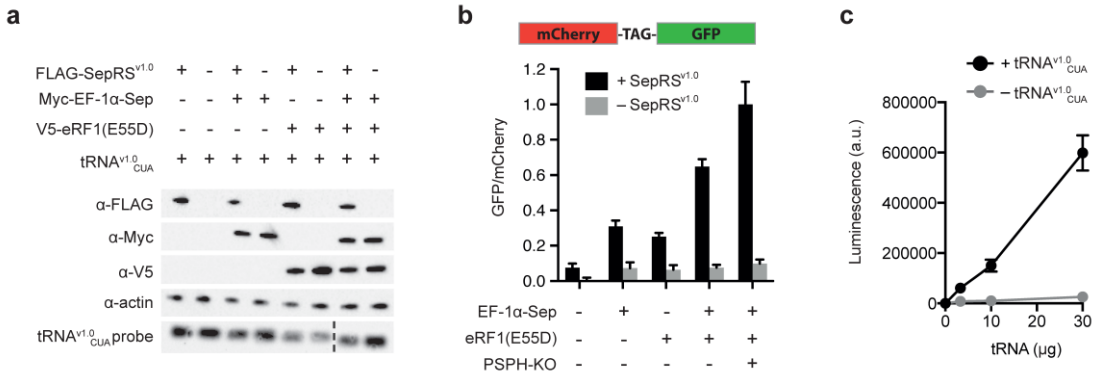

Figure 2.1. The SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ pair enables pSer incorporation in mammalian cells. (**a**) Expression of FLAG-SepRS$^{v1.0}$, Myc-EF-1α-Sep, V5-eRF1(E55D) and tRNA$^{v1.0}_{CUA}$ in HEK293 cells detected via western and northern blots. Dashed line indicates removal of irrelevant lanes. (**b**) SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ pair directs pSer incorporation in mammalian cells, co-expression of eEF1-αSep, eRF1(E55D) and knockout of PSPH gene increase pSer incorporation. Readthrough of UAG codon is determined by ratio of GFP to mCherry fluorescence from the mCherry-TAG-GFP reporter. Data represent mean ± SEM for at least three biological replicates. (**c**) Recombinant SepRS$^{v1.0}$ is incubated with total tRNA extracted from control mammalian cells (– tRNA$^{v1.0}_{CUA}$) or cells expressing tRNA$^{v1.0}_{CUA}$ (+ tRNA$^{v1.0}_{CUA}$). The aminoacylation in each reaction is determined by measuring the AMP production with the AMP-Glo assay. (Data represent mean ± SEM for triplicate reactions.) (Data in panel a were obtained by C.D.R., data in panel b were obtained by V.B. and data in panel c were obtained by M.S.Z.. C.D.R. created the PSPH knockout cell line.)

To assess the functionality of the SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ pair in human cells we measured the

readthrough of the amber stop codon in an mCherry-TAG-GFP construct via the ratio of GFP and mCherry fluorescence. Expression of $tRNA^{v1.0}_{CUA}$ alone led to minimal read through of the amber stop codon, consistent with this tRNA being minimally aminoacylated by endogenous synthetases. This observation suggests that $tRNA^{v1.0}_{CUA}$ is orthogonal in mammalian cells (**Figure 2.1b**). Addition of SepRS led to an increase in amber suppression consistent with the SepRS mediated aminoacylation of $tRNA^{v1.0}_{CUA}$ with phosphoserine, which is present in mammalian cells as a biosynthetic precursor to serine (Ichihara and Greenberg, 1957).

In *E. coli*, the incorporation of phosphoserine into proteins by the $SepRS^{v1.0}/tRNA^{v1.0}_{CUA}$ pair can be enhanced by a mutant form of EF-Tu, named EF-Sep[8], which may accommodate the phosphorylated amino acid. EF-1α is the eukaryotic equivalent of EF-Tu, and delivers aminoacylated tRNAs to the eukaryotic ribosome (Carvalho et al., 1984). We therefore asked whether SepRS dependent readthrough of the amber stop codon in mammalian cells can be enhanced by a mutant of EF-1α. To design a mutant of EF-1α for pSer incorporation in mammalian cells we aligned the EF-Tu and EF-1α sequences and identified the amino acids in EF-Tu that were mutated to create EF-Sep (**Figures 2.2a and 2.2b**). We then created the corresponding mutations (L77R, Q251N, D252G, V264S, N307W) in EF-1α to create EF-1α-Sep, and demonstrated that this mutant was expressed in mammalian cells (**Figure 2.1a**). Co-expression of EF-1α-Sep with the $SepRS^{v1.0}/tRNA^{v1.0}_{CUA}$ pair led to an increase in the read through of the amber stop codon in mCherry-TAG-GFP (**Figure 2.1b**).

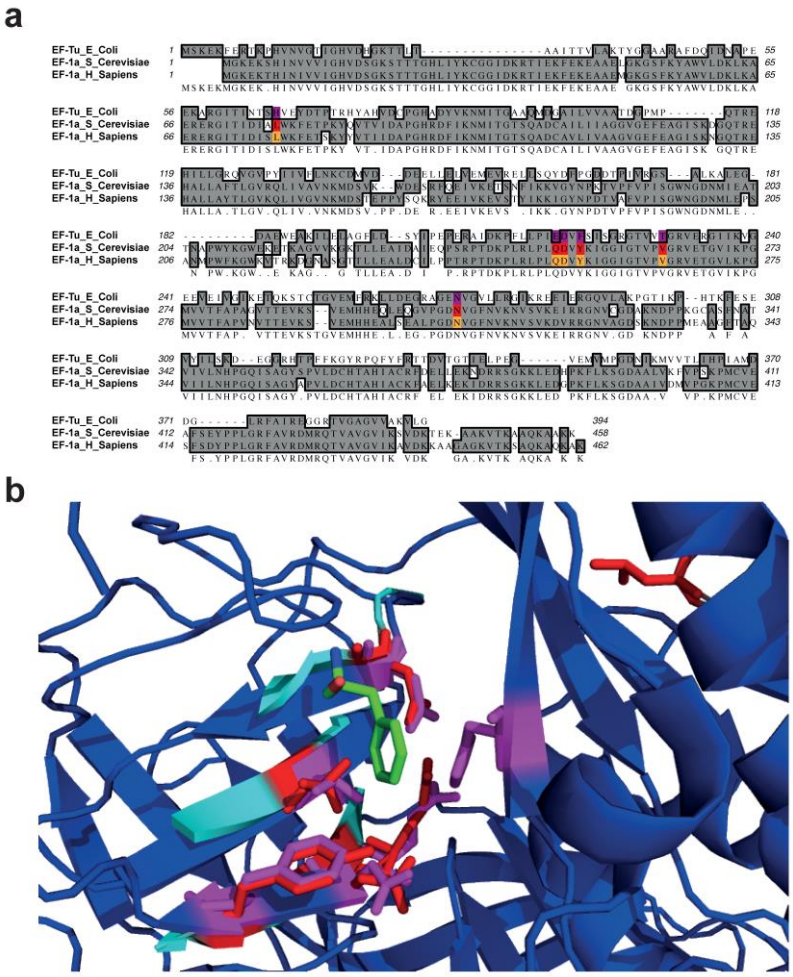

Figure 2.2. Generation of EF-1α-Sep. (a) Clustal alignment of protein sequences of EF-Tu (E. coli) and EF-1α (*S. cerevisiae* and *H. sapiens*). Residues mutated to create EF-Sep (Park et al., 2011) (H67R, E216N, D217G, F219Y, T229S, N274W) are highlighted in magenta, corresponding residues in EF-1α are highlighted in red (*S. cerevisiae*) and orange (H. sapiens). **(b)** Aligned structures of *E. coli* EF-Tu in complex with phenylalanyl-tRNA (PDB:1ob2; EF-Tu in blue, Phe in green, tRNA not shown) and *S. cerevisiae* EF-1α (PDB:1f60; cyan, only structure surrounding the critical residues shown). Mutated residues are highlighted in magenta (*E. coli*) and red (*S. cerevisiae*). (Figure created by V.B.).

We have previously shown that the efficiency of ncAA incorporation, using the PylRS/$tRNA_{CUA}$ pair, in mammalian cells can be enhanced by expressing an eRF1(E55D) mutant of eRF1 (Schmied et al, 2014; Kolosov et al., 2005); the eukaryotic release factor that terminates protein synthesis at all three stop codons. eRF1(E55D) does not efficiently terminate protein synthesis in response to amber codons, and its expression minimizes release factor

competition with amber suppressor tRNAs at the amber stop codon, while maintaining termination on the other two stop codons. Co-expression of eRF1(E55D) with the SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ pair was confirmed by immunoblotting (**Figure 2.1a**) and led to a substantial increase in read through of the amber stop codon (**Figure 2.1b**).

Next, we combined the SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ with both EF-1α-Sep and eRF1(E55D). This combination led to the greatest read through of the amber codon in the presence of SepRS$^{v1.0}$ (**Figure 2.1b**). These data are consistent with a model in which SepRS$^{v1.0}$, which is known to selectively recognize phosphoserine (Hauenstein et al., 2008), aminoacylates the orthogonal tRNA$^{v1.0}_{CUA}$ with phosphoserine in mammalian cells, creating pSer-tRNA$^{v1.0}_{CUA}$. pSer-tRNA$^{v1.0}_{CUA}$ is taken to the ribosome with the aid of EF-1α-Sep, where it is more efficiently decoded in response to the amber codon in the presence of eRF1(E55D). All further experiments were carried out using the SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ with EF-1α-Sep and eRF1(E55D). To verify that transfection of the SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$, EF-1α-Sep and eRF1(E55D) does not compromise cellular fitness, we compared the cell viability with cells transfected with control plasmid (**Figure 2.3**) and saw no significant difference. We note that both the altered translation factors (EF-1α-Sep and eRF1(E55D)) mediate some SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ pair independent readthrough of the amber stop codon; this may result from near-cognate decoding of the amber codon by endogenous tRNAs (Roy et al., 2015).

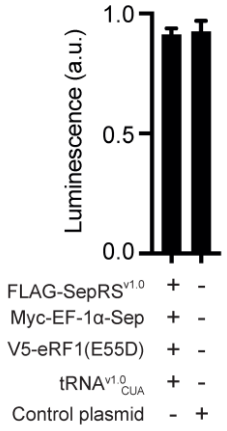

Figure 2.3. CellTiter-Glo assay confirms that cell viability is not compromised by transfection of the genetic code expansion components when compared to transfection of a control plasmid. (Data measured by M.S.Z.).

2.2.2.2 PSPH deletion increases intracellular pSer

Does the intracellular concentration of phosphoserine limit the efficiency of its incorporation using our system in mammalian cells? To address this question, we extracted metabolites from cells, and found that the phosphoserine levels were less than 100 μM *in vivo* (**Figure 2.4**). Since the *Km* of SepRS for phosphoserine is approximately 270 μM (Hauenstein et al., 2008) we reasoned that increasing the pSer concentration in cells might increase the efficiency of its incorporation into proteins. In mammals, phosphoserine phosphatase (PSPH) converts phosphoserine to serine in the last step of serine biosynthesis (Snell, 1984) and we hypothesized that knocking out PSPH might lead to an increase in intracellular phosphoserine levels and allow us to test the effect of phosphoserine levels on SepRS mediated incorporation into proteins. We performed CRISPR-Cas9 mediated knockout of PSPH in HEK293, and confirmed the knockout by genotyping and western blot (**Figure 2.5**). In the resulting cell line, HEK293/PSPH-KO, the intracellular pSer concentration increased by at least 400 ± 60 μM (SD) over HEK293 (**Figure 2.6**). This increase in intracellular phosphoserine led to a measureable increase in phosphoserine incorporation in response to the amber codon in the HEK293/PSPH-KO (**Figure 2.1b**). We conclude that phosphoserine levels in mammalian cells can be increased by PSPH deletion. Overall, the use of EF-1α-Sep, eRF1(E55D) and the PSPH knockout increase the efficiency of $SepRS^{v1.0}/tRNA^{v1.0}_{CUA}$ mediated amber suppression by more than an order of magnitude.

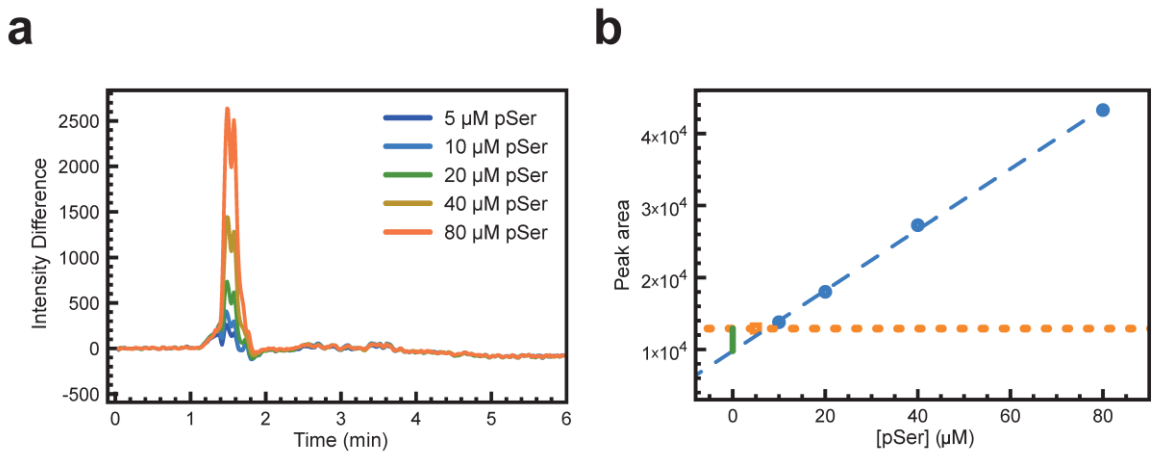

Figure 2.4. Intracellular concentration of phosphoserine in HEK293. **(a)** The LC-MS chromatograms obtained from subtracting the unspiked HEK293 lysates from each of the pSer-spiked samples are graphed as a function of time. **(b)** The peak area of each pSer-spiked sample plotted as a function of the spiked pSer concentration. The original chromatograms were used for analysis rather than the difference data. The dashed, blue line represents a linear fit from 10-80 µM pSer. The dotted orange line represents the peak area for the 5 µM pSer standard. The green line represents the difference between the peak area for 5 µM pSer and the projected peak area for a sample with no pSer spiked in, corresponding to approx. 100 µM pSer. (Data measured by A.D.L.).

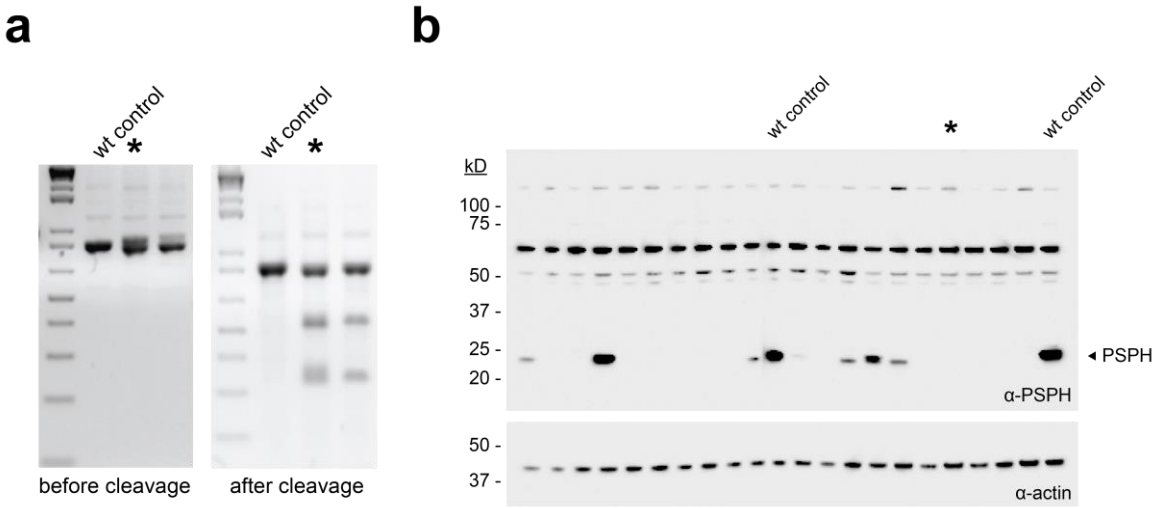

Figure 2.5. Characterization of HEK293/PSPH-KO. **(a)** Gel electrophoresis of PCR amplicons of PSPH exon 2 from HEK293/PSPH-KO clonal cell lines before (left) and after reannealing and treatment with Guide-it resolvase (right). Star denotes the final clone used in this study. **(b)** Confirmation of PSPH knockout using western blot. Star denotes the final clone used in this study. (Data measured by C.D.R.).

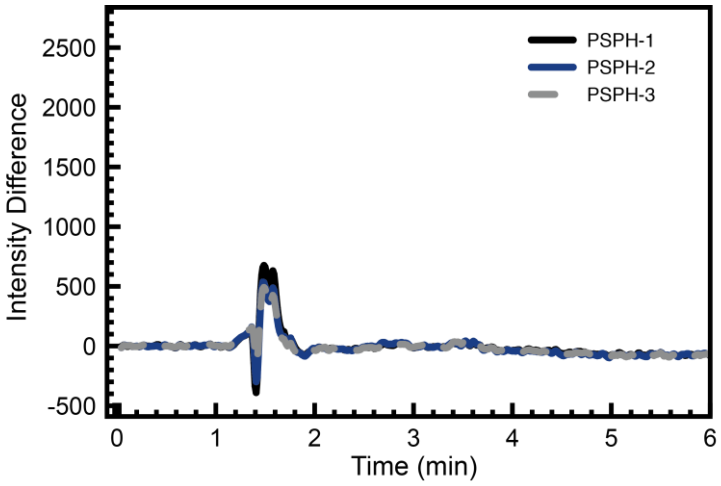

Figure 2.6. Intracellular concentration of phosphoserine in HEK293/PSAT-KO. The LC-MS chromatograms obtained from subtracting the unspiked HEK293 lysates from each of the HEK293/PSPH-KO lysate samples. Peak areas correspond to increase in pSer concentration of 400 ± 60 μM (SD) over HEK293. (Data measured by A.D.L.).

2.2.2.3 SepRS$^{v1.0}$ is orthogonal with respect to mammalian tRNA

Next we demonstrated that SepRS$^{v1.0}$ is selective for tRNA$^{v1.0}_{CUA}$ with respect to the mammalian tRNAs. We isolated total tRNA from HEK293 cells (-tRNA$^{v1.0}_{CUA}$) and from HEK293 cells expressing tRNA$^{v1.0}_{CUA}$ (+tRNA$^{v1.0}_{CUA}$), in which tRNA$^{v1.0}_{CUA}$ makes up less than 10% of the total mammalian tRNA pool (**Figure 2.7**). We subjected each tRNA pool to *in vitro* aminoacylation with phosphoserine using purified SepRS$^{v1.0}$. We followed the extent of aminoacylation as a function of total tRNA concentration by measuring AMP production (Mondal et al., 2017). For +tRNA$^{v1.0}_{CUA}$ we observed an increase in aminoacylation with total tRNA concentration, while for - tRNA$^{v1.0}_{CUA}$ we observed minimal aminoacylation at all tRNA concentrations tested (**Figure 2.1c**). Our results demonstrate that SepRS$^{v1.0}$ does not substantially aminoacylate endogenous mammalian tRNAs but selectively aminoacylates tRNA$^{v1.0}_{CUA}$. We conclude that SepRS$^{v1.0}$ is orthogonal with respect to the tRNAs in mammalian cells.

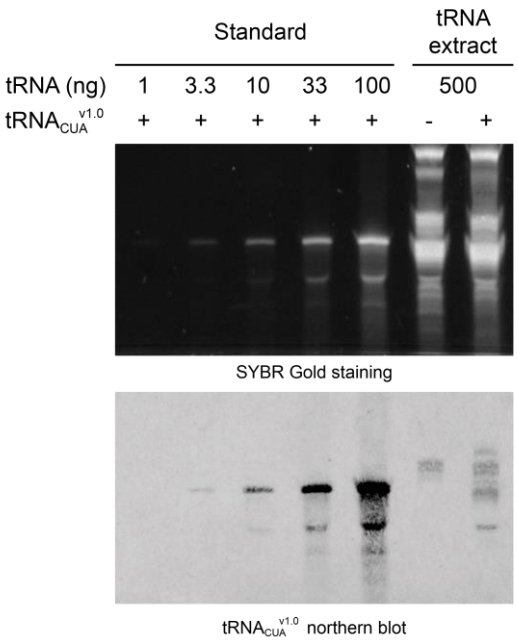

Figure 2.7. Northern blot confirms presence of tRNA$^{v1.0}_{CUA}$ in mammalian tRNA extract. SYBR Gold staining of RNA after denaturing PAGE (top panel). Lanes 1 to 5 contain fixed amounts of a chemically synthesised RNA oligo with the same sequence of tRNA$^{v1.0}_{CUA}$ in the range from 1 to 100 ng. The last two lanes contain a total of 500 ng of RNA extract (measured by UV absorbance) from wild type HEK293 cells, or from cells transfected with plasmid containing 4x tRNA$^{v1.0}_{CUA}$ gene under U6 promoter. Northern blot using a fluorescently labelled probe specific for tRNA$^{v1.0}_{CUA}$ (bottom panel). No specific signal can be detected from the RNA extract of wild type cells. The signal detectable from the tRNA$^{v1.0}_{CUA}$ expressing cells is slightly greater than the one measured from the standard containing 10 ng of RNA, but lower than the one from the standard containing 33 ng of RNA. (Data measured by M.S.Z).

2.2.2.4 Encoded pSer is post-translationally converted to Ser

To investigate the identity of the amino acid incorporated into proteins in response to the amber codon we created a streamlined expression system in which SepRS, eRF1(E55D), EF-1α-Sep and four copies of tRNA$^{v1.0}_{CUA}$ are combined on a single plasmid. Co-transfection of this plasmid with a plasmid containing GFP(150TAG) and four copies of tRNA$^{v1.0}_{CUA}$ into HEK293 cells enabled expression and purification of the resulting GFP (**Figure 2.8a**).

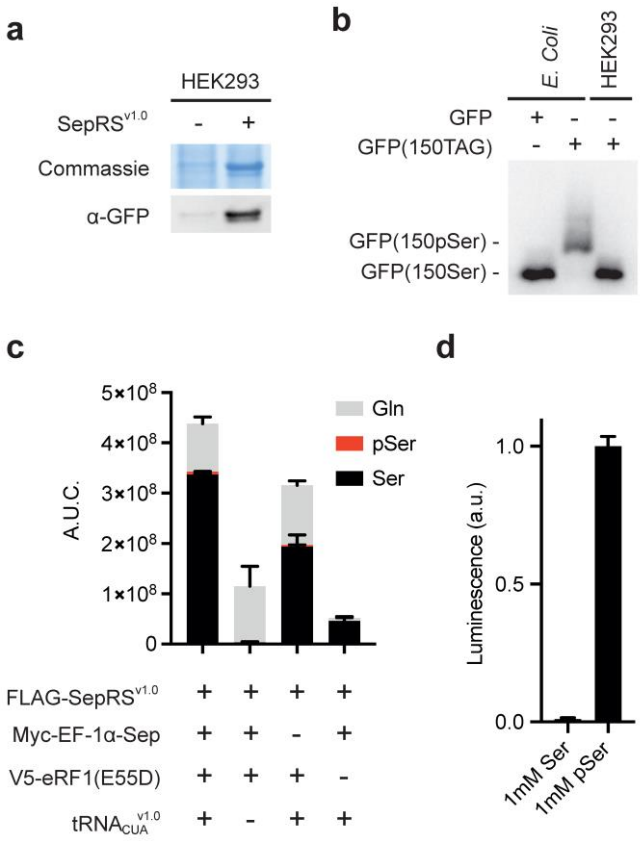

Figure 2.8. SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ directs pSer into proteins, where pSer is post-translationally dephosphorylated. **(a)** Coomassie-stained SDS-PAGE gel and western blot of purified GFP from HEK293 cells. **(b)** pSer is not maintained post-translationally in GFP expressed in mammalian cells. The Phos-tag SDS-PAGE gel leads to a mobility shift in phosphorylated proteins via chelation of the phosphate in the gel. GFP and GFP(150pSer) standards were produced in *E. coli* as described previously (Rogerson et al., 2015), and define the mobility of phosphorylated and non phosphorylated GFP. GFP was detected by immunoblotting. **(c)** A.U.C. is the area under the curve of the extracted ion chromatograms for peptide LEYNFNSH[X]VYITADK in MS1 of the tryptic LC-MS/MS. Identity of the peptides was confirmed by MS/MS analysis (see **Figure 2.9**). **(d)** SepRS$^{v1.0}$ is selective for phosphoserine over serine. Recombinant SepRS$^{v1.0}$ was incubated with total tRNA extracted from mammalian cells expressing tRNA$^{v1.0}_{CUA}$. The aminoacylation in each reaction was determined by measuring the AMP production with AMP-Glo assay. Data represent mean ± SEM for reactions in four replicates. (Data in panels a, b and c measured by V.B., data in panel d measured by M.S.Z.).

To investigate whether phosphoserine was present in GFP expressed in mammalian cells we prepared authentic standards of GFP bearing phosphoserine at position 150 (GFP(150pSer)), using a previously described *E. coli* expression system (Rogerson et al. 2015), and GFP bearing

serine at position 150 (GFP(150Ser)). We analyzed the samples on phos-tag-SDS-PAGE, a well-established technique in which phosphorylated proteins are resolved from non-phosphorylated proteins and the extent of phosphorylation can be quantified (Kinoshita et al., 2006; Kinoshita et al., 2009).

As expected, the mobility of GFP(150pSer) was retarded with respect to that of GFP(150Ser) (**Figure 2.8b**). The majority of GFP expressed from GFP(150TAG) in mammalian cells migrated with the GFP150Ser standard prepared in *E. coli*, and we were not able to detect any phosphorylated GFP. This result demonstrates that the majority of the amino acid at position 150 in GFP, expressed from GFP(150TAG) in mammalian cells, is not phosphorylated.

To investigate the identity of amino acids incorporated in response to the amber codon in GFP we performed electrospray ionization tandem mass spectrometry (LC-MS/MS) on tryptic digests of GFP expressed in mammalian cells. The extracted ion chromatographs of peptides at the MS1 level reveal that serine, glutamine and traces of phosphoserine are found in GFP produced from GFP(150TAG) in mammalian cells containing SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$, EF-1α-Sep and eRF1(E55D) (**Figure 2.8c**). Control experiments show that the relevant peptides containing pSer and Ser ionize with comparable efficiency (**Figures 2.9a and 2.9b**).

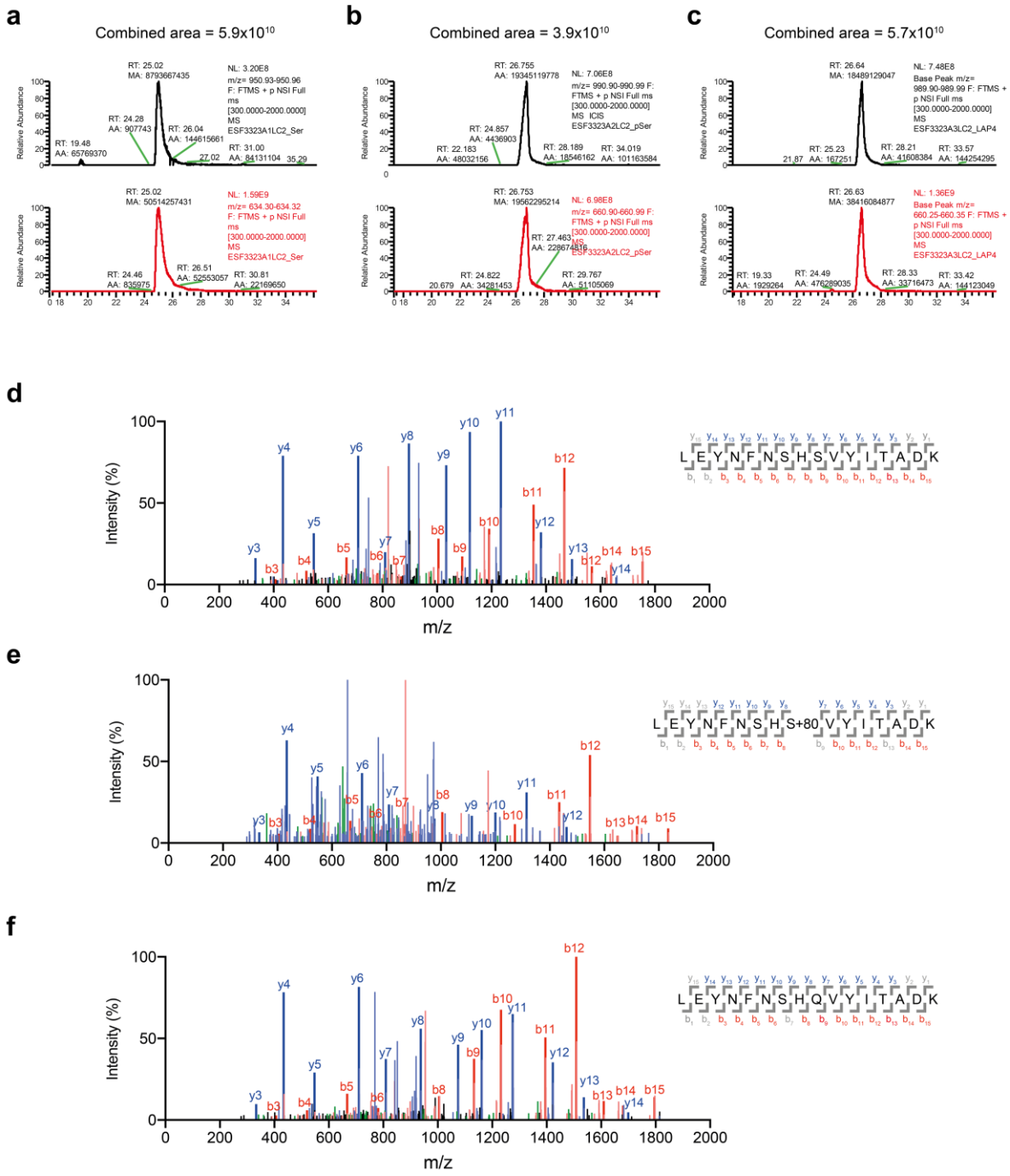

Figure 2.9. LC-MS quantification and LC-MS/MS identification of serine, phosphoserine and **2 (Figure 2.10a)** incorporation using E. coli standards. Related to **Figure 2.10**. Extracted ion chromatograms of 2+ (black line) and 3+ (red line) ions of the tryptic peptides LEYNFNSH[X]VYITADK where **(a)** [X] = serine, **(b)** [X] = phosphoserine and **(c)** [X] = 2 when the instrument was injected with 1:1:1 mixture of GFP150Ser, GFP150pSer and GFP1502. Combined areas under the curves for 2+ and 3+ ions (shown in figures) demonstrate that the ionization of the three peptides is comparable. MS/MS spectrum confirms incorporation of serine **(d)**, phosphoserine **(e)** and glutamine **(f)** at position 150 in

sfGFP150TAG extracted from mammalian cells. We detected no other amino acids at the position 150. (Data in panels measured by V.B.).

We suspected that the glutamine incorporation we observe in response to the amber codon in mammalian cells may result from suppressor tRNA independent incorporation of glutamine upon expressing eRF1(E55D) in mammalian cells (Roy et al. 2015), consistent with the $SepRS^{v1.0}$ independent readthrough of the amber codon observed with eRF1(E55D) in our initial system (**Figure 2.1b**). Protein expression in the absence of eRF1(E55D) led to a decrease in GFP synthesis (**Figure 2.8c**), as expected based on the role of this mutant in decreasing the efficiency of translational termination at the amber stop codon. It also led to a substantial increase in the ratio of serine to glutamine at position 150 of GFP (**Figure 2.8c**), consistent with a decrease in non-cognate decoding of the TAG codon when the normal termination machinery is active. In contrast, protein expression in the absence of EF-Sep led to a decrease in protein, and an increase in the ratio of glutamine to serine at position 150 (**Figure 2.8c**), consistent with this factor increasing the delivery of phosphoseryl-$tRNA^{v1.0}_{CUA}$ to the ribosome and post translational dephosphorylation of phosphoserine. Finally, protein expression in the absence of the amber suppression machinery (-$tRNA^{v1.0}_{CUA}$) but in the presence of eRF1(E55D) led to a decrease in GFP, and the protein predominantly incorporated glutamine at position 150 (**Figure 2.8c**), indicating that the glutamine incorporation is not mediated by the $SepRS^{v1.0}$/$tRNA^{v1.0}_{CUA}$ pair. These observations are consistent with previous reports of glutamine incorporation in response to premature termination codons in eukaryotic mRNAs (Roy et al., 2015).

The observed serine incorporation could formally result from either the co-translational incorporation of serine by $SepRS^{v1.0}$/$tRNA^{v1.0}_{CUA}$ or the incorporation of phosphoserine followed by its posttranslational dephosphorylation by phosphatases in cells. SepRS is known to be highly selective for phosphoserine over serine. Previous work has shown that SepRS selectively forms the aminoacyl-adenylate of phosphoserine over that of serine by a factor of 10^4 in the first step of aminoacylation (Hauenstein et al., 2008), and the overall kinetics of aminoacylation of serine with this system were too low to measure; these observations suggest that SepRS discriminates against serine by much more than 10^4 fold in the overall aminoacylation reaction (Hauenstein et al., 2008). Consistent with these prior observations,

the aminoacylation of $tRNA^{v1.0}_{CUA}$ with 1mM serine by $SepRS^{v1.0}$ was negligible (**Figure 2.8d**). In contrast, $tRNA^{v1.0}_{CUA}$ was efficiently aminoacylated by $SepRS^{v1.0}$ in the presence of 1 mM phosphoserine. These observations further confirm that $tRNA^{v1.0}_{CUA}$ is not aminoacylated with serine by $SepRS^{v1.0}$.

Our observations suggest that in mammalian cells our system directs the co-translational incorporation of phosphoserine in response to the amber codon, and that most of the genetically encoded phosphorylation in GFP is post-translationally dephosphorylated to serine. As numerous phosphatases for serine and threonine residues are present in mammalian cells (Hornbeck et al., 2004) this observation is not surprising.

2.2.2.5 Encoding a non-hydrolyzable pSer

Next we asked whether we could increase the fraction of phosphorylated protein by encoding a non-hydrolyzable phosphonate analog of pSer (**2**) (**Figure 2.10a**) in cells depleted in phosphoserine and fed **2**. We first established that **2** can enter mammalian cells and is a good substrate for SepRS. Supplementing the growth media with 2 mM and 10 mM **2** led to intracellular concentrations of 3.64 ± 0.39 mM and 12.47 ± 0.80 mM, respectively (**Figure 2.11**) demonstrating that we can generate millimolar concentrations of amino acid **2** in the cytosol. *In vitro* aminoacylation experiments confirm that $SepRS^{v1.0}$ can efficiently aminoacylate $tRNA^{v1.0}_{CUA}$ with **2** (**Figure 2.12**), in line with our previous observations incorporating the amino acid in *E. coli* (Rogerson et al., 2015).

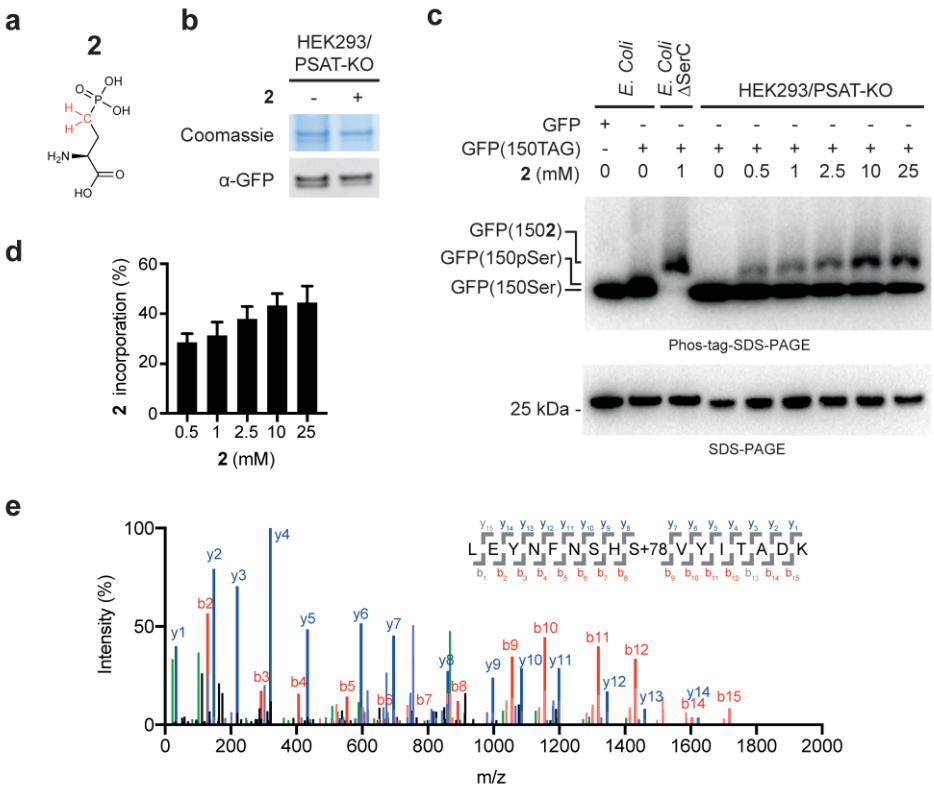

Figure 2.10. Encoding non-hydrolyzable phosphonate analogue (2) of pSer in genetically engineered mammalian cells. **(a)** Phosphonate analogue of phosphoserine (**2**) used in this study. **(b)** Protein expression in the PSAT-KO. Coomassie-stained SDS-PAGE gel and western blot of purified GFP from HEK293/PSAT-KO overexpressing PSPH. Amino acid **2** was used at 10 mM **(c)** Separation on phos-tag SDS-PAGE gel followed by immunoblotting is consistent with incorporation of **2** in the HEK293/PSAT-KO cell line overexpressing PSPH. GFP, GFP(150pSer) and GFP(150(**2**)) standards were produced in *E. coli* as described previously (Rogerson et al., 2015). **(d)** Quantification of the relative incorporation of **2** as a result of increasing concentration of **2** added to the cells. Data represent mean ± SEM for three biological replicates. **(e)** Incorporation of **2** into GFP(150TAG) reporter at the genetically encoded site was verified by ESI-MS/MS. See also **Figure 2.9**. (Data measured by V.B., HEK293/PSAT-KO cell line was created by C.D.R.).

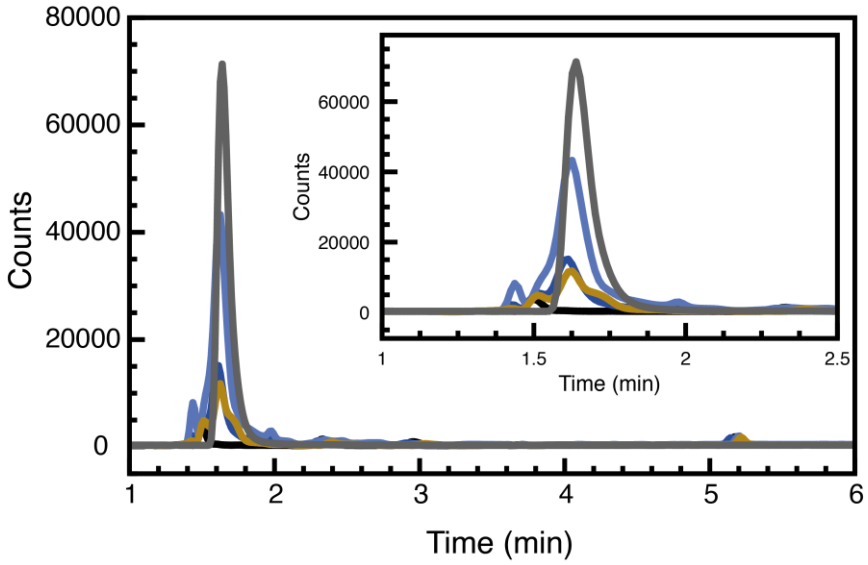

Figure 2.11. Phosphonate analogue is taken up by HEK293 cells from culture media. The LC-MS chromatograms of lysates from wtHEK293 cultured in standard growth media (black), wtHEK293 supplemented with 2 mM **2** in growth media (dark blue), wtHEK293 supplemented with 20 mM **2** in growth media (light blue), wtHEK293 with 500 μM **2** added to the lysate (final concentration) (gold) and aqueous standard of 2 (grey). Peak areas correspond to intracellular concentration of 3.64 ± 0.39 mM and 12.47 ± 0.80 mM intracellular **2** for cells cultured in media supplemented with 2 mM and 20 mM **2**, respectively. All values represent mean ± SD from three biological replicates. The insert is magnified section of the chromatograms. (Data measured by A.D.L.).

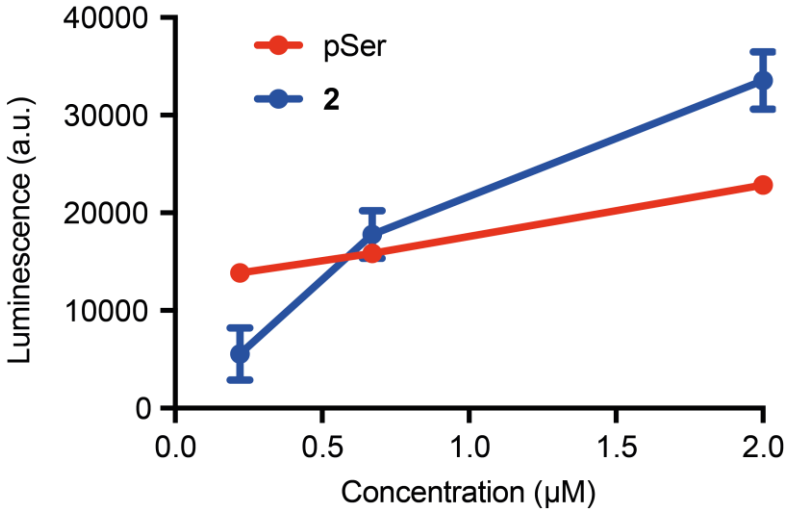

Figure 2.12. 2 is a substrate for SepRS$^{v1.0}$ in vitro. Recombinant SepRS$^{v1.0}$ was incubated with total tRNA extracted from cells transfected with tRNACUADTR and the reactions were supplemented with phosphoserine or **2** in varying concentrations. The aminoacylation in each

reaction was determined by measuring the AMP production with an AMP-Glo assay. The luminescence of reactions was normalised by subtracting the luminescence of reaction with no amino acid. Data represent mean ± SEM for three replicates. (Data measured by M.S.Z.).

Since pSer is biosynthesized from phosphohydroxypyruvate via the action of PSAT and converted to serine via the action of PSPH (Ichihara and Greenberg, 1957), we reasoned that phosphoserine levels resulting from this pathway could be minimized by both the deletion of PSAT and the overexpression of PSPH. We created a CRISPR-Cas9 mediated knockout cell line, HEK293/PSAT-KO, and confirmed this knockout by genotyping and western blot (**Figure 2.13**). pSer levels were already below the detection limit of our LC-MS assay in HEK293 cell metabolite extractions, and we could not measure these levels in the HEK293/PSAT-KO strain. We found that the effect of the PSAT knockout and PSPH overexpression on serine levels was modest, reducing them from 362 ± 52 μM in HEK293 cells to 201 ± 22 μM and 215 ± 14 μM (all errors are SD) for the PSAT knockout and PSAT knockout with PSPH overexpression, respectively (**Figure 2.14**).

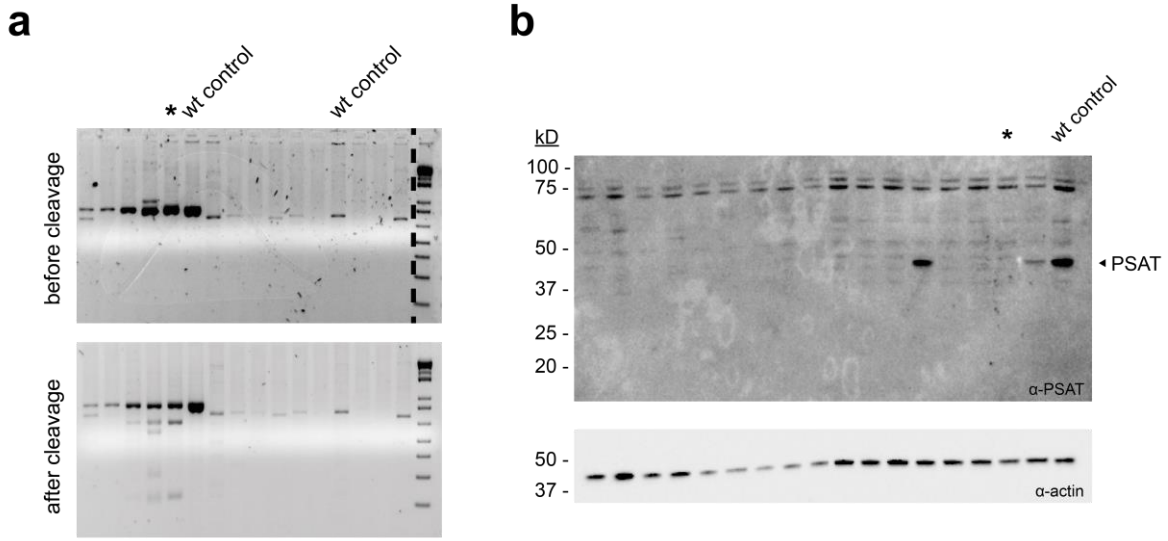

Figure 2.13. Characterization of HEK293/PSPH-KO and HEK293/PSAT-KO cell lines. (**a**) Gel electrophoresis of PCR amplicons of PSAT exon 1 from HEK293/PSAT-KO clonal cell lines before (top) and after reannealing and treatment with Guide-it resolvase (bottom). Star denotes the clone used in this study. Dashed line indicates unrelated lanes that were removed for clarity. (**b**) Confirmation of PSAT knockout using western blot. Star denotes the clone used

in this study. (Both cell lines created by C.D.R. Data measured by C.D.R.).

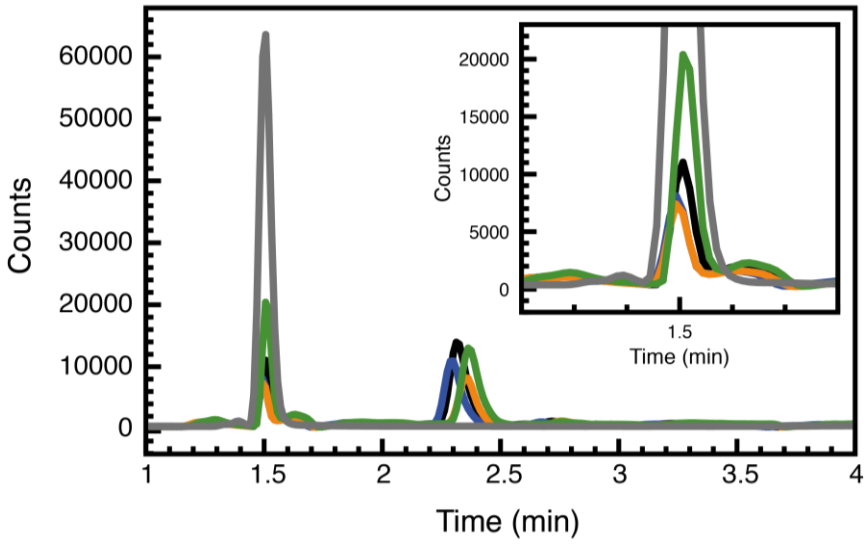

Figure 2.14. The LC-MS chromatograms of lysates from wtHEK293 (black), HEK293/PSAT-KO (dark blue), HEK293/PSAT-KO with overexpressed PSPH (orange), wtHEK293 with spiked 50 μM Serine (final concentration) (green) and aqueous standard for Serine (grey). Peak areas correspond to intracellular Serine concentration of 362 ± 52 μM, 201 ± 22 μM, 215 ± 14 μM for wtHEK293, HEK293/PSAT-KO and HEK293/PSAT-KO with PSPH overexpression, respectively. All values represent mean ± SD from three biological replicates. The insert is magnified section of the chromatograms. (Data measured by A.D.L.).

We co-transfected HEK293/PSAT-KO cells with two plasmids: one encoding SepRS$^{v1.0}$, EF-1α-Sep, eRF1(E55D), PSPH and four copies of tRNA$^{v1.0}_{CUA}$, and one encoding GFP(150TAG) and four copies of tRNA$^{v1.0}_{CUA}$. GFP was expressed from these cells via read-through of the amber codon. Protein production in these cells was comparable in the presence and absence of **2** (**Figure 2.10b**).

To investigate whether our system enables **2** to compete with phosphoserine for SepRS$^{v1.0}$ mediated incorporation into proteins in mammalian cells we took advantage of our ability to prepare proteins that site-specifically incorporate either phosphoserine or **2** in *E. coli*.[9] We produced authentic standards of GFP incorporating pSer at position 150 (GFP(150pSer)) and GFP incorporating **2** at position 150 (GFP(150**2**)) in *E. coli* (Figure S2a-c), and subjected these

(along with GFP encoding serine at position 150 (GFP(150Ser))) to phos-tag-SDS-PAGE (**Figure 2.10c**).

We discovered that the phosphonate in **2** leads to a dramatic decrease in mobility of GFP(150**2**), with respect to both GFP and GFP(150pSer), in phos-tag-SDS-PAGE (we ran these gels under conditions where GFP150Ser and GFP150pSer are not resolved, as longer runs led to diffuse bands of GFP150**2**). This result demonstrates that GFP incorporating **2** can be resolved by its mobility in phos-tag-SDS-PAGE (**Figure 2.10c**). GFP expressed from GFP(150TAG) in mammalian cells deleted for PSAT, and expressing PSPH, $SepRS^{v1.0}$, EF-1α-Sep, eRF1(E55D) and $tRNA^{v1.0}_{CUA}$, and provided with **2** led to two bands on phos-tag-SDS-PAGE; one band co-migrated with the authentic standard for GFP(150**2**) and the other with the authentic standards for GFP(150Ser) and GFP150pSer (**Figure 2.10c**, top panel). Standard SDS-PAGE analysis of the same sample produced single band for each sample, confirming that the slower migration of the upper band in phos-tag SDS-PAGE is due to a different interaction of **2** with the phos-tag gel matrix (**Figure 2.10c**). The GFP(150**2**) fraction of the total GFP produced increased with the concentration of **2** in the growth media and reached 49 ± 11% (SEM) for 25 mM **2**. Mass spectrometry (**Figure 2.10e** and **Figure 2.15**) revealed that the GFP produced contained either glutamine or **2**; this suggests that **2** can outcompete pSer at the active site of SepRS (control experiments demonstrate that the relevant peptides containing pSer and **2** ionize comparably (see **Figure 2.9**)), but that the amber suppressor tRNA does not entirely outcompete glutamine incorporation in the presence of eRF1(E55D). We note that the ratio of **2** to glutamine in these experiments is lower than the ratio of serine to glutamine when phophoserine is incorporated (**Figure 2.8c**); this may be a result of the different genetic backgrounds used for the two experiments.

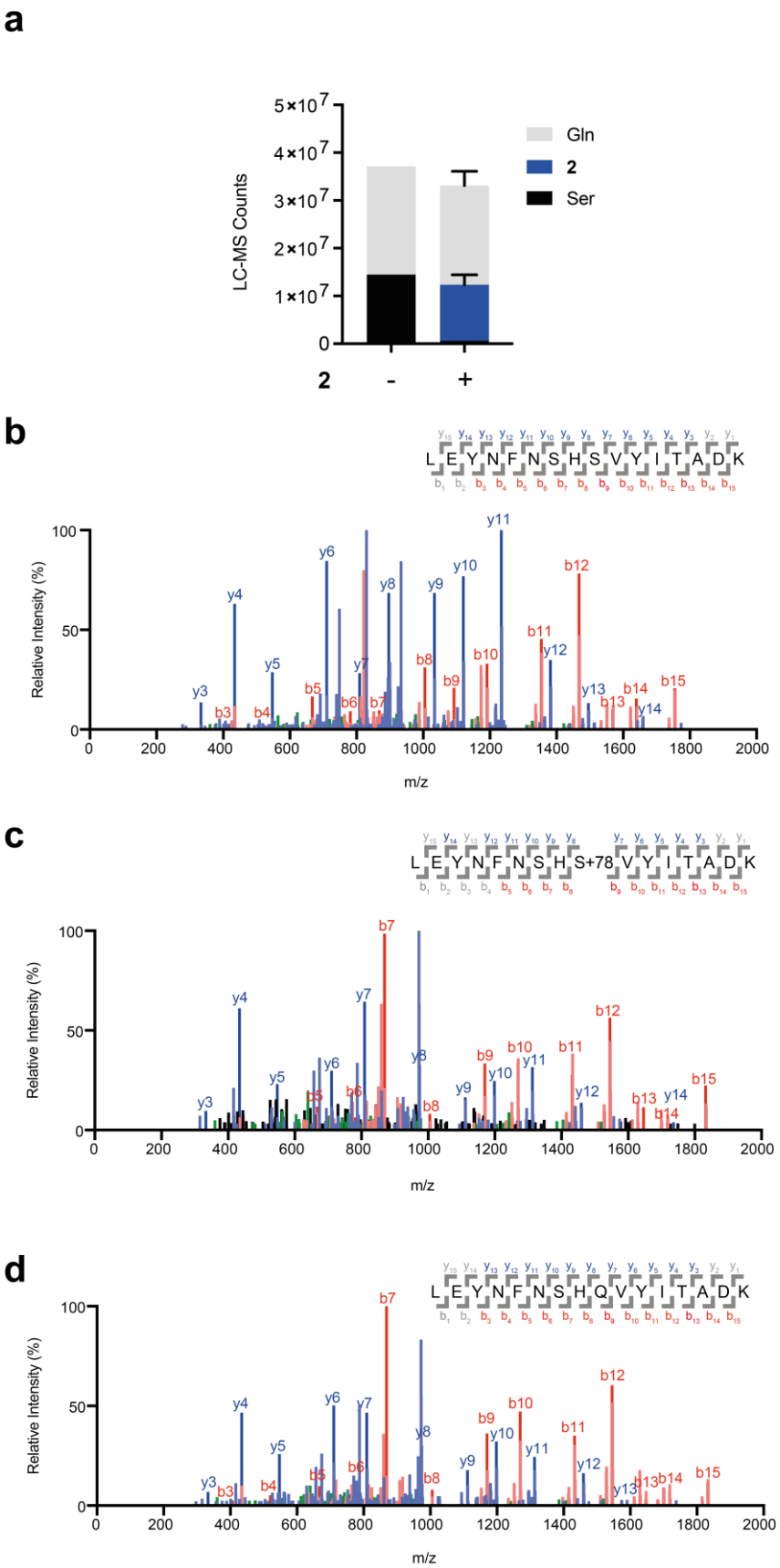
a
LC-MS Counts
Gln
2
Ser
2
-
+
b
L E Y N F N S H S V Y I T A D K
Relative Intensity (%)
m/z
c
L E Y N F N S H S+78 V Y I T A D K
Relative Intensity (%)
m/z
d
L E Y N F N S H Q V Y I T A D K
Relative Intensity (%)
m/z

Figure 2.15. Identification of amino acid incorporated in response to amber codon in sfGFP150TAG. **(a)** LC-MS counts of tryptic peptides containing serine, glutamine or **2** at position 150 in sfGFP (no other amino acids were detected). Data are represented as mean ± SEM, where relevant. MS/MS spectrum confirms incorporation of serine **(b)**, 2 **(c)** and glutamine **(d)** at position 150. No other amino acids were detected. (Data measured by V.B.).

Since phosphorylation is often an activating, dominant modification and glutamine incorporation in place of serine is very unlikely to have phenotypic consequences (and these can be explicitly ruled out by testing the phenotype of an Ser to Gln mutation). Moreover, since proteomics studies suggest that the majority of proteins are naturally regulated by sub-stoichiometric phosphorylation (<30%) (Olsen et al., 2010), the sub-stoichiometric genetic encoding of phosphorylated amino acids in mammalian cells should not substantively limit the utility of our system for studies in mammalian cells.

2.2.2.6 Kinase activation by genetically encoded phosphorylation

Next we explicitly demonstrated that genetically encoding **2** at a biologically relevant site of serine phosphorylation enables synthetic activation of a protein kinase (MEK1) in mammalian cells. It is well established that phosphorylation at serine 218 and serine 222 of MEK1, upon stimulation of upstream kinases by growth factors, leads to activation of MEK1 and phosphorylation of its substrates, notably ERK1/2 (Shaul and Seger, 2007). Serine to aspartic acid substitutions at positions 218 and 222 can activate MEK1 (Huang and Erikson, 1994). MEK1 is also phosphorylated at several other sites (Hornbeck et al., 2004) and the native protein from unstimulated cells runs as two bands in a phostag gel (**Figure 2.16**), reflecting the phosphorylation state of un-activated MEK1 in mammalian cells.

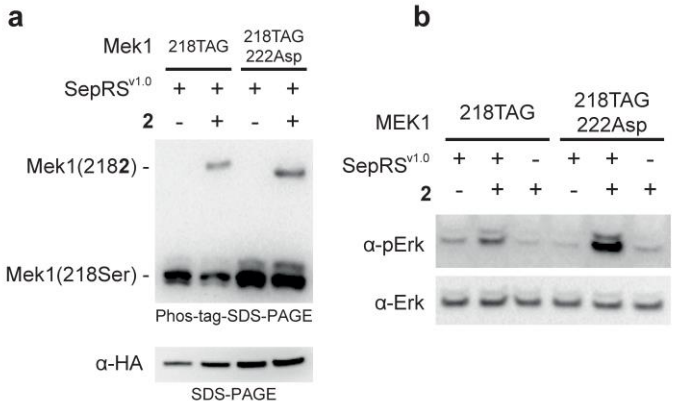

Figure 2.16. Activation of Mek by incorporation of 2. **(a)** Phos-tag SDS PAGE analysis is consistent with incorporation of **2** at position 218 in Mek1/2 (top panel). SDS-PAGE analysis (bottom panel) shows single band, confirming that the slower migration of upper band is due to chelation of phosphate group in the gel. **(b)** Expression of MEK1(218**2**) and MEK1(218**2**222D) results in phosphorylation of Erk in SepRS$^{v1.0}$ dependent manner, as confirmed by phosphospecific antibody (top panel). The loading control was performed with standard Erk antibody (bottom panel). (Data measured by V.B.).

We introduced plasmids expressing MEK1(218TAG) or MEK1(218TAG222Asp), along with an (ERK)-GFP fusion, SepRS$^{v1.0}$, EF-1α-Sep, eRF1(E55D), tRNA$^{v1.0}_{CUA}$ and PSPH into HEK293/PSAT-KO cells. As expected, based on the GFP expression experiments, addition of **2** led to stably phosphorylated MEK variants, which were retarded in Phos-tag SDS PAGE (Figure 4a) with respect to the bands characteristic of unactivated MEK.

Incorporation of **2** at position 218 of MEK led to phosphorylation of ERK-GFP (Figure 4b), consistent with reports that a serine to aspartic acid mutation at this position can stimulate some kinase activity (Huang et al., 1995). We observed a more dramatic increase in ERK-GFP phosphorylation by MEK(218**2**222D), consistent with maximal activity of MEK requiring both phosphorylations (Huang et al., 1995). Control experiments in which either SepRS$^{v1.0}$ or **2** are omitted substantially decreased levels of ERK-GFP phosphorylation, further confirming that the activation we observe results from the genetically encoded incorporation of **2**.

2.2.3 Discussion

We have demonstrated that the SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ pair is orthogonal in mammalian cells. We have developed an EF-1α mutant that facilitates the incorporation of phosphoserine, and combined this with engineered eRF1 and manipulations of phosphoserine biosynthesis to increase the efficiency of phosphoserine incorporation. By minimizing the intracellular pool of phosphoserine and feeding the cell a non-hydrolyzable analog of phosphoserine we have been able to genetically encode a constitutive phosphorylation in mammalian cells. We anticipate that the first generation system we have reported herein may be further optimized by directed evolution of the translational components in mammalian cells. In this regard we note that the current PylRS/tRNA$_{CUA}$ systems are orders of magnitude more efficient than the original systems (Schmied et al., 2014).

The demonstration that the SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ pair is orthogonal in mammalian cells and can be combined with other factors for the incorporation of pSer in mammalian cells, along with our recent demonstration that the SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ can be evolved in *E. coli* to recognize ncAAs that are not natural substrates of SepRS (Zhang et al., 2017) opens up many new opportunities. We note that it may be possible to evolve the SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ to incorporate ncAAs in *E. coli* and to use the evolved SepRS$^{v1.0}$ variants to incorporate diverse ncAAs into proteins in mammalian cells.

We used our approach to activate MEK1 by co-translationally encoding a phosphorylated amino acid at genetically targeted sites *in vivo*. We anticipate that extensions of this approach may provide a route to deciphering the consequences of specific phosphorylations, both through synthetically manipulating cellular signaling pathways, as demonstrated here, and as a basis for capturing phospho-protein interaction partners in cells.

2.2.4 References

Burgess, S.G., Mukherjee, M., Sabir, S., Joseph, N., Gutiérrez- Caballero, C., Richards, M.W., Huguenin- Dezot, N., Chin, J.W., Kennedy, E.J., and Pfuhl, M. (2018). Mitotic spindle association of TACC3 requires Aurora- A- dependent stabilization of a cryptic α- helix. The EMBO journal, e97902.

Carvalho, M.D.G.D.C., Carvalho, J.F., and Merrick, W.C. (1984). Biological characterization of various forms of elongation factor 1 from rabbit reticulocytes. Archives of Biochemistry and Biophysics 234, 603-611.

Chin, J.W. (2014). Expanding and reprogramming the genetic code of cells and animals. Annual review of biochemistry 83, 379-408.

Chin, J.W. (2017). Expanding and reprogramming the genetic code. Nature 550, 53.

David, Y., Vila-Perelló, M., Verma, S., and Muir, T.W. (2015). Chemical tagging and customizing of cellular chromatin states using ultrafast trans-splicing inteins. Nature chemistry 7, 394-402.

Davis, L., and Chin, J.W. (2012). Designer proteins: applications of genetic code expansion in cell biology. Nature reviews Molecular cell biology 13, 168-182.

Dickson, C., Fletcher, A.J., Vaysburd, M., Yang, J.-C., Mallery, D.L., Zeng, J., Johnson, C.M., McLaughlin, S.H., Skehel, M., and Maslen, S. (2018). Intracellular antibody signalling is regulated by phosphorylation of the Fc receptor TRIM21. eLife 7, e32660.

Elsässer, S.J., Ernst, R.J., Walker, O.S., and Chin, J.W. (2016). Genetic code expansion in stable cell lines enales encoded chromatin modification. Nature methods 13, 158-164.

Hauenstein, S.I., Hou, Y.-M., and Perona, J.J. (2008). The homotetrameric phosphoseryl-tRNA synthetase from Methanosarcina mazei exhibits half-of-the-sites activity. Journal of Biological Chemistry 283, 21997-22006.

Hornbeck, P.V., Chabra, I., Kornhauser, J.M., Skrzypek, E., and Zhang, B. (2004). PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. Proteomics 4, 1551-1561.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., and Shalem, O. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature biotechnology 31, 827-832.

Huang, W., and Erikson, R.L. (1994). Constitutive activation of Mek1 by mutation of serine phosphorylation sites. Proceedings of the National Academy of Sciences 91, 8960-8963.

Huang, W., Kessler, D., and Erikson, R. (1995). Biochemical and biological analysis of Mek1 phosphorylation site mutants. Molecular biology of the cell 6, 237-245.

Huguenin-Dezot, N., De Cesare, V., Peltier, J., Knebel, A., Kristaryianto, Y.A., Rogerson, D.T., Kulathu, Y., Trost, M., and Chin, J.W. (2016). Synthesis of isomeric phosphoubiquitin chains reveals that phosphorylation controls deubiquitinase activity and specificity. Cell reports 16, 1180-1193.

Ichihara, A., and Greenberg, D.M. (1957). Further studies on the pathway of serine formation from carbohydrate. Journal of Biological Chemistry 224, 331-340.

Kinoshita, E., Kinoshita-Kikuta, E., and Koike, T. (2009). Separation and detection of large phosphoproteins using Phos-tag SDS-PAGE. Nature protocols 4, 1513-1521.

Kinoshita, E., Kinoshita-Kikuta, E., Takiyama, K., and Koike, T. (2006). Phosphate-binding tag, a new tool to visualize phosphorylated proteins. Molecular & Cellular Proteomics 5, 749-757.

Kolosov, P., Frolova, L., Seit-Nebi, A., Dubovaya, V., Kononenko, A., Oparina, N., Justesen, J., Efimov, A., and Kisselev, L. (2005). Invariant amino acids essential for decoding function of polypeptide release factor eRF1. Nucleic acids research 33, 6418-6425.

Mandell, D.J., Chorny, I., Groban, E.S., Wong, S.E., Levine, E., Rapp, C.S., and Jacobson, M.P. (2007). Strengths of hydrogen bonds involving phosphorylated amino acid side chains. Journal of the American Chemical Society 129, 820-827.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science 298, 1912-1934.

Mondal, S., Hsiao, K., and Goueli, S.A. (2017). Utility of Adenosine Monophosphate Detection System for Monitoring the Activities of Diverse Enzyme Reactions. Assay and drug development technologies 15, 330-341.

Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127, 635-648.

Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., and Nigg, E.A. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal 3, ra3-ra3.

Park, H.-S., Hohn, M.J., Umehara, T., Guo, L.-T., Osborne, E.M., Benner, J., Noren, C.J., Rinehart, J., and Söll, D. (2011). Expanding the genetic code of Escherichia coli with phosphoserine. Science 333, 1151-1154.

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T.C., and Waldo, G.S. (2006). Engineering and characterization of a superfolder green fluorescent protein. Nature biotechnology 24, 79-88.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature protocols 8, 2281-2308.

Roberts-Galbraith, R.H., Ohi, M.D., Ballif, B.A., Chen, J.-S., McLeod, I., McDonald, W.H., Gygi, S.P., Yates, J.R., and Gould, K.L. (2010). Dephosphorylation of F-BAR protein Cdc15 modulates its conformation and stimulates its scaffolding activity at the cell division site. Molecular cell 39, 86-99.

Rogerson, D.T., Sachdeva, A., Wang, K., Haq, T., Kazlauskaite, A., Hancock, S.M., Huguenin-Dezot, N., Muqit, M.M., Fry, A.M., and Bayliss, R. (2015). Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. Nature chemical biology 11, 496-503.

Roy, B., Leszyk, J.D., Mangus, D.A., and Jacobson, A. (2015). Nonsense suppression by near-cognate tRNAs employs alternative base pairing at codon positions 1 and 3. Proceedings of the National Academy of Sciences 112, 3038-3043.

Sauerwald, A., Zhu, W., Major, T.A., Roy, H., Palioura, S., Jahn, D., Whitman, W.B., Yates, J.R., Ibba, M., and Söll, D. (2005). RNA-dependent cysteine biosynthesis in archaea. Science 307, 1969-1972.

Schmied, W.H., Elsässer, S.J., Uttamapinant, C., and Chin, J.W. (2014). Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1. Journal of the American Chemical Society 136, 15577-15583.

Shaul, Y.D., and Seger, R. (2007). The MEK/ERK cascade: from signaling specificity to diverse functions. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research 1773, 1213-1226.

Snell, K. (1984). Enzymes of serine metabolism in normal, developing and neoplastic rat tissues. Advances in enzyme regulation 22, 325-400.

Tsai, Y.-H., Essig, S., James, J.R., Lang, K., and Chin, J.W. (2015). Selective, rapid and optically switchable regulation of protein function in live mammalian cells. Nature chemistry 7, 554-561.

Zhang, M.S., Brunner, S.F., Huguenin-Dezot, N., Liang, A.D., Schmied, W.H., Rogerson, D.T., and Chin, J.W. (2017). Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. Nature Methods 14, 729-736.

2.2.5 METHODS

2.2.5.1 EXPERIMENTAL MODEL AND SUBJECT DETAILS

HEK293 cells (epithelial human embryonic kidney, female)(ATCC) were cultivated in DMEM (Gibco) supplemented with 10% fetal bovine serum (Gibco) in humifdified incubator at 37°C, 5% CO_2. For passaging, cells were washed with phosphate buffered saline, detached using trypsin/EDTA-solution, resuspended in fresh growth medium and seeded into cell culture flasks. The cells were routinely tested for mycoplasma contamination.

2.2.5.2 METHOD DETAILS

2.2.5.2.1 Plasmid Generation

$tRNA^{v1.0}_{CUA}$ (B4 variant (Rogerson et al., 2015)) repeats containing the human U6 promoter, followed by the $tRNA^{v1.0}_{CUA}$ with no CCA tail, and a terminator were synthesized with unique homology arms on each side, allowing for cloning of four repeats by Gibson into a minimal backbone derived from pUC19. The resulting sequence of one repeat is as following:

Human U6 promoter :: $\underline{tRNA^{v1.0}_{CUA}}$ :: terminator

CCTAGTTGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATA
CGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAG
ATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAG
TTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC<u>GCCGGGGTAG
TCTAGGGGTTAGGCAGCAGTCTCTAAAATTGCCTTACGTGGGTTCAAATCCCACC
CCCGGCT</u>GACAAGTGCGGTTTTT

The structure of EF-1α was aligned with EF-Sep, and the engineered mutations were transferred over (see Figures S1a and S1b). The mutations (L77R, Q251N, D252G, V264S, N307W) were introduced via QuickChange site-directed mutagenesis and the resulting EF-1α-Sep was cloned into pcDNA3.4 via Gibson. $SepRS^{v1.0}$ was cloned into pcDNA3.4 via Gibson from previously published plasmids (Rogerson et al., 2015).

Plasmids containing four copies of $tRNA^{v1.0}_{CUA}$, protein of interest ($SepRS^{v1.0}$, EF-1α-Sep, or multiple cloning site (MCS)) under EF1A promoter were cloned by Gibson assembly from the above described plasmids and previously published plasmids (Schmied et al, 2014). The resulting constructs are designated pPB_FLAG-$SepRS^{v1.0}$_4xU6-$tRNA^{v1.0}_{CUA}$, pPB_Myc-EF-1α-SepRS_4xU6-$tRNA^{v1.0}_{CUA}$ and pPB_MCS_4xU6-$tRNA^{v1.0}_{CUA}$. Fluorescence reporter constructs pPB_GFP(150TAG)_4xU6-$tRNA^{v1.0}_{CUA}$ (GFP refers to sfGFP variant (Pédelacq et al., 2006)) with C-terminal His_6 tag) and pPB_mCherry-TAG-GFP_4xU6-$tRNA^{v1.0}_{CUA}$ were assembled from previously reported plasmids (Elsässer et al., 2016) via restriction cloning into pPB_MCS_4xU6-$tRNA^{v1.0}_{CUA}$. pPB_V5-eRF1(E55D)_4xU6-$tRNA^{v1.0}_{CUA}$ was similarly cloned via restriction cloning into pPB_MCS_4xU6-$tRNA^{v1.0}_{CUA}$. Control plasmids with no $tRNA^{v1.0}_{CUA}$ (pPB_MCS and pPB_mCherry-TAG-GFP) were cloned using Gibson assembly from the plasmids described above. pPB_FLAG-SepRS_Myc-EF1a_V5-eRF1(E55D)_4xU6-$tRNA^{v1.0}_{CUA}$ and pPB_FLAG-SepRS_Myc-EF1a_V5-eRF1(E55D)_V5-PSPH_4xU6-$tRNA^{v1.0}_{CUA}$ were enzymatically assembled via Gibson from plasmids described above with individual coding DNA sequences joined via cleavable P2A peptides. pPB_MEK1(218TAG)_4xU6-$tRNA^{v1.0}_{CUA}$ and pPB_MEK1(218TAG222Asp)_4xU6-$tRNA^{v1.0}_{CUA}$ were cloned by gibson assembly from previously reported plasmids (Tsai et al., 2015).

2.2.5.2.2 Phosphoserine Incorporation

HEK293 cells were seeded into poly-L-lysine coated 24 well plates (1.4×10^5 cells per well). Upon reaching 70%-80% confluency cells were transfected with combinations of pPB_FLAG-$SepRS^{v1.0}$_4xU6-$tRNA^{v1.0}_{CUA}$, pPB_Myc-EF-1α-Sep _4xU6-$tRNA^{v1.0}_{CUA}$, pPB_V5-eRF1(E55D)_4xU6-$tRNA^{v1.0}_{CUA}$ and pPB_mCherry-TAG-GFP_4xU6-$tRNA^{v1.0}_{CUA}$. For data in Figure 1, we transfected 125 ng of each plasmid per well, keeping the total amount of transfected DNA at 500 ng by addition of pPB_MCS_4xU6-$tRNA^{v1.0}_{CUA}$ or pPB_MCS. The DNA was transfected using 2 µL of 1 mg/ml solution of polyethylenimine MW 40,000 (PEI; Polysciences) per well. For subsequent experiments, the cells were transfected with combination of pPB_FLAG-SepRS_Myc-EF1a_V5-eRF1(E55D)_V5-PSPH_4xU6-$tRNA^{v1.0}_{CUA}$ or SepRS_Myc-EF1a_V5-eRF1(E55D)_V5-PSPH_4xU6-$tRNA^{v1.0}_{CUA}$ and

pPB_GFP(150TAG)_4xU6-tRNA$^{v1.0}_{CUA}$ using 1000 ng of DNA/well. DNA and PEI were diluted in Opti-MEM and incubated for 15 minutes at room temperature. Cells were washed once with Opti-MEM before adding the transfection mix to the cells. The transfection medium was exchanged for fresh growth medium after 4 hours. The medium was exchanged again two days after the transfection. Cells were harvested and analyzed after 3 to 4 days of expression. Cell viability was measured using CellTiter-Glo 2.0 assay (Promega) according to the manufacturer's instructions.

Flow Cytometry

The cells were washed with PBS, detached using trypsin/EDTA solution, resuspended in growth medium, pelleted and resuspended in PBS. The cells were analyzed using Becton Dickinson LSRII SORP (488 nm coherent sapphire laser for GFP excitation, 561 coherent compass laser for mCherry excitation). The data was analyzed in FlowJo software (FlowJo, LLC).

Western Blotting

For lysis HEK293 cells were washed with PBS and lysed in RIPA buffer containing HALT Protease and Phosphatase Inhibitor Cocktail (Thermo) and incubated at 4°C for 5 minutes. Samples were centrifuged for 10 minutes at 20000 rcf. Protein concentration of the supernatant was determined via BCA assay (ThermoFisher) according to manufacturer's instructions. The sample was transferred to SDS sample buffer (62.5 mM Tris-HCl (pH 7.4), 12.6% glycerol, 2% SDS, 100 mM DTT, 0.01% bromophenol blue) and boiled at 95°C for 5 minutes. Proteins were separated via SDS-PAGE using 4-12% Bis-Tris gels (Novex) and transferred to nitrocellulose membranes using the iBlot2 Dry Blotting System (ThermoFisher) according to manufacturer's instructions.

The membrane was blocked with 5% milk in tris-buffered saline with 0.1% Tween 20 (TBS-T) for 1 h and incubated with the primary antibody α-GFP (Roche (13.1 and 7.1)), α-FLAG (Sigma-Aldrich (F1804)), α-Myc (Cell Signaling (9B11)), α-V5 (Invitrogen (2F11F7)), α-actin (Sigma-Aldrich (A3853)), α-HA (Roche (3F10)), α-Erk (Cell Signalling (L34F12)), α-pErk (Cell Signalling (D13.14.4E)) at 4 °C overnight. The membrane was washed several times with TBS-T and incubated with corresponding secondary antibody conjugated to HRP (Santa Cruz

Biotechnology or Cell Signalling) for 1 h at room temperature. The blot was washed with TBS-T and developed using SuperSignal West Pico or Femto Chemiluminescent Substrate (ThermoFisher) according to manufacturer's instructions. Images were acquired using a ChemiDoc system (Bio-Rad).

Northern Blotting

RNA was isolated from HEK293 cells using the QIAzol Lysis Reagent (QIAGEN) according to manufacturers instructions. The samples were analyzed using the NorthernMax-Gly procedure (Thermo Fisher). Sep-tRNA was detected using a biotin labeled probe (AACCCACGTAAGGCAATTTTAGAGACTGCTGCCTAACC-CCT), which was hybridized at 37 °C overnight in Ambion ULTRAhyb Ultrasensitive Hypridization Buffer (ThermoFisher). Blots were washed with low stringency wash buffer (2x saline-sodium citrate, 0.1% sodium dodecyl sulfate), incubated with horseradish peroxidase conjugated streptavidin and developed using enhanced chemiluminescence solution (both ThermoFisher).

2.2.5.2.3 Cell Line Generation

To knock out PSPH in HEK293 cells a guide RNA targeting exon 2 of PSPH was designed using to previously reported software (Hsu et al., 2013) and cloned into the pSpCas9(BB)-2A-GFP plasmid using a previously reported protocol (Ran et al., 2013) (top strand: CACCGCACACCTGACCCCCGGCATA; bottom strand: AAACTATGCCGGGGGTCAGGTGTGC). HEK293 cells were transfected with the plasmid using PEI and grown in DMEM containing 10% FBS and 1 mM serine for 3 days before they were sorted to single cells via fluorescence activated cell sorting based on GFP fluorescence. Subsequently, single cell colonies were screened for indel formation using the Guide-it mutation detection kit (Clontech laboratories) and a complete PSPH knock out was confirmed by western blot analysis using α-PSPH antibody (ThermoFisher). The PSAT knockout HEK293 cell line (HEK293/PSAT-KO) was created analogously using guide RNA targeting exon 1 of PSAT (top strand: CACCGCCAGGCAGGTGGTCAACTTT; bottom strand: AAACAAAGTTGACCACCTGCCTGGC). The primers used for genotyping via the Guide-it mutation detection kit were: CAGCCTGTTAATGTTATTTTCAAGC (PSPH exon 2 forward primer), ACGCTGTAGTAAGCCATGATTATAC (PSPH exon 2 reverse primer),

AAACAGTAAACGCGAGGAGG (PSAT exon 1 forward primer), CTCATTCACACTATGTCCATTCATGC (PSAT exon 1 reverse primer).

2.2.5.2.4 In Vitro Experiments

SepRS Purification

The gene fragment coding for SepRS was cloned into a pET-20b vector to give pET20b_SepRS. BL21(DE3) cells transformed with pET20b_SepRS were used to inoculate 2-litre 2xTY culture. At OD600 = 0.5, 0.1 mM IPTG was added and the culture was incubated at 24 °C for 16hrs before being pelleted. The cell pellet was resuspended in lysis buffer (50 mM Tris-HCl, pH = 8.0, 300 mM NaCl, 20 mM imidazole, 1 mM beta-mercaptoethanol). After sonication, the lysates were heated to 70°C for 20 min and centrifuged at 20 000g for 30 min. The supernatant was loaded onto a HisTrap column which was eluted by a linear gradient between 20 mM to 500 mM imidazole. The fractions with SepRS were pooled and incubated with 3C protease at 4 °C for 24hrs to remove the tag. Then, proteins were further loaded to HiTrap Q column and eluted with a gradient from 50 mM to 1 M NaCl. The fractions with SepRS were pooled, concentrated, and snap frozen in liquid nitrogen and stored at -80°C.

tRNA Extraction and Purification

HEK293 cells were transfected as described above with plasmid containing 4xU6-$tRNA^{v1.0}_{CUA}$ cultured for 2 days and harvested by pipetting into ice cold PBS. 500 µl of cell pellet are resuspended in 6 ml of tRNA Buffer (50 mM NaOAc pH 5, 150 mM NaCl, 10 mM MgCl2, 0.1 mM EDTA) and aliquoted into 5x2 ml vial. The cells are centrifuged again (5 min, 500 rcf), the supernatant removed and 950 µl of fresh tRNA buffer are added. To lyse the cells, 50 µl of liquefied phenol (P9346, Sigma Aldrich) are added and the tubes put in mild head-over-tail shaking (10 rpm) for 15 min. The vials are then centrifuged for 25 min at 4 °C at 21,000 rcf, then the supernatant containing the tRNAs is recovered and transferred to a new vial, where 1 volume of chloroform is added. After thorough mixing using a vortex, the emulsion is separated in a centrifuge (1 min, 21,000 rcf) and the acqueous phase recovered. To deacylate the purified tRNAs, 40 µl of NaOH 300 mM are added to 680 µl of aqueous phase. The reaction is incubated for 1 h at 42 °C, then the solution is neutralised using 40 µl of NaOAc 3 M pH 5. The deacylated tRNAs are ethanol precipitated for 1 h at 10 °C by adding 2.33 volumes of

absolute ethanol to the recovered aqueous phase. After 30 min centrifugation at 21,000 rcf at 4 °C the supernatant is removed and the precipitated RNAs resuspended in water.

In Vitro Aminoacylation Assay

The aminoacylation by SepRS was measured with AMP-Glo assay (Promega). 25 nM recombinant SepRS was mixed with 0, 3.3, 10, or 30 µg extracted tRNA in aminoacylation buffer (50 mM Tris pH = 7.4, 20 mM KCl, 2 mM DTT, 10 mM MgCl2) supplemented with 0.1 mM ATP, 1 mM phosphoserine, and 0.1 mg/ml BSA. Reactions were assembled to 25 µl and were incubated at 37 °C for 2 hrs. The AMP generated by the reaction was then measured with AMP-Glo assay by following the instruction from the manufacturer.

2.2.5.2.5 Quantitative Northern Blot Analysis

500 ng of the RNAs purified from wild type cells or cell transfected with a plasmid containing the $tRNA^{v1.0}_{CUA}$ are run alongside linear dilutions of RNA standard on a 10% acrylamide gel under denaturing conditions (7 M urea, 1xTBE buffer) for 70 min at 200 V using the BioRad Mini-proean apparatus. The gel is then stained using SYBR Gold and imaged under blue light to visualise total RNA, then the RNA is transferred on positively charged nylon membrane using iBolt transfer system and crosslinked to the membrane under UV light. The membrane was pre-hypridised using the ULTRAhyb®-Oligo Buffer for 30 min at 37°C, then 2 µg of infrared fluorescent probe [5'-/5IDR800/GTG GGA TTT GAA CCC ACG TAA GGC AAT TTT -3', Integrated DNA Technologies] were added. The hybridisation was carried out at 37°C overnight. The excess probe was washed using 2xSSC buffer containing 0.1% SDS four times, then the membrane imaged using the Amersham Typhoon 5 (785 nm laser).

2.2.5.2.6 LC-MS Analysis of Intracellular Amino Acids

HEK293, HEK293/PSPH-KO and HEK293/PSAT-KO cells were cultured and transfected as described above. Cells were detached using trypsin and washed three times with PBS before analysis. Concentration was measured using Countess II cell counter (ThermoFisher) prior to last pelleting of the cells. The frozen cell pellets were thawed at room temperature and resuspended in methanol-water (40:60) to 500 µl. The suspension was transferred to a 1.5-ml

eppendorf tube. The cells were lysed by six freeze-thaw cycles (liquid nitrogen and room temperature). The resulting lysate was centrifuged at 21,000 g for 1 h at 4 ˚C. The supernatant was pipetted into a 3K MWCO amicon centrifugal filter. The samples were centrifuged at 14,000 g for 30 min at 4 ˚C. The flow-throw was analysed by LC-MS analysis. From the flow-through, 50 µl aliquots were pipetted into 250-µl glass inserts (Agilent). To prepare a calibration curve, 50-µl aliquots of the control sample were spiked with serine, pSer, or **2** to a final concentration of 5, 10, 20, 40, or 80 µM. For comparison, analogous calibration curves were also prepared in a methanol-water mixture (40:60).

An Agilent 1260 Infinity equipped with an Agilent 6130 Quadrupole LC-MS unit was used for analysis of all samples. From each sample, 10 µl was injected onto a Zorbax SB C18 column, 4.6 x 150 mm equipped with a guard column (Agilent). Each sample was eluted from the column using a mobile phase gradient from 0.5 % to 95 % acetonitrile containing 0.02 % formic acid. The mass spectrometer was set to selected ion monitoring (SIM) mode. For pSer, the ions measured were 186 M/z in the positive mode and 184 M/z in the negative mode. For Ser, the ions measured were 106 M/z in the positive mode and 104 M/z in the negative mode. Lastly, for **2**, the ions measured were 184 M/z in the positive mode and 182 M/z in the negative mode. The peaks for phosphoserine and serine overlapped with background peaks. Thus difference spectra were prepared for analysis by subtracting the chromatogram of the control with the chromatograms from each calibration standard. The resulting traces were integrated to derive a calibration curve. For the pSer lysate samples, the chromatograms from the HEK293 samples were subtracted from the chromatograms of the HEK293/PSPH-KO cell line. From the difference chromatograms, the unobscured peak for phosphoserine was integrated. A linear fit of the calibration curve from 10-80 µM was used to determine the concentration of serine, phosphoserine, and **2** in each of the lysates.

Using the following equation, the intracellular concentration was determined from the lysate concentration and the cell number measurements for each sample. [IC] = (lysate concentration × lysate volume) / (total number of cells × cell volume). The lysate volumes were 500 µl for all experiments. An approximate value was used for the cell volume, 2 pL. The intracellular concentration of phosphoserine in the control HEK cells was roughly estimated by extrapolation of the calibration curve and comparison with the 5 µM calibration point.

2.2.5.2.7 Purification of GFP from Mammalian Cells

HEK293 or HEK23/PSAT-KO cells were seeded into T-75 flasks (5.25×10^6 per flask) 24 hours before transfection with 18.75 μg of DNA (75 μL PEI). The medium was changed after 4 hours to medium with or without 10 mM **2**, neutralized with NaOH. Cells were harvested, lysed and GFP was purified via GFP pull-down using GFP-Trap_MA beads (ChromoTek GmbH) according to manufacturer's instructions 3 days after transfection. Proteins were eluted by addition of SDS-sample buffer and boiling at 95 °C for 10 minutes. Alternatively, proteins were eluted using 200 mM glycine buffer (pH 2), which was neutralized by addition of Tris HCl (pH 10.4). The proteins were separated via SDS-PAGE, stained using InstantBlue (Expedeon).

2.2.5.2.8 Phos-tag SDS-PAGE Electrophoresis

Cells were cultured and transfected as described above. Before harvesting, cells were washed twice with TBS and detached using pipette or cell scraper. The cells were lysed in RIPA buffer supplemented with HALT protease and phosphatase inhibitors (ThermoFischer). Protein concentration was measured using BCA assay (ThermoFischer) according to manufacturer's instructions. 8% Phos-tag SDS-PAGE gels (Wako) were run according to manufacturer's instructions (25 mA, 100-120 minutes) and the proteins were transferred onto PVDF membrane using iBlot2 system (ThermoFischer). The membrane was immunostained and developed as described above. GFP, GFP150pSer and GFP150(**2**) standards were expressed in *E. Coli* as described previously (Rogerson et al., 2015). Quantitative analysis was done using ImageJ.

2.2.5.2.9 LC-MS/MS Analysis of Amino Acid Incorporation

The tandem mass spectrometry was carried out as described previously (Rogerson et al., 2015). Briefly, bands of interest (1–2 mm) were cut from polyacrylamide gel. The gel slices were destained with 50% v/v acetonitrile and 50 mM ammonium bicarbonate, reduced with 10 mM DTT and alkylated with 55 mM iodoacetamide in 96-well plate. The proteins were digested overnight at 37 °C using 6 ng/μl trypsin solution (Promega) and the resulting peptides were extracted with 2% v/v formic acid, 2% v/v acetonitrile. The peptides were analyzed by

nanoscale capillary LC-MS/MS using Ultimate U3000 HPLC (ThermoScientific Dionex) under 300 nl/min flow. Before separation, the peptides were trapped on a C18 Acclaim PepMap100 3 μm, 75 μm × 150 mm nanoViper (ThermoScientfic Dionex) and subsequently eluted with acetonitrile gradient into a modified nanoflow ESI source with a hybrid dual pressure linear ion trap mass spectrometer (Orbitrap Velos, ThermoScientific). Analysis was carried out using a resolution of 30,000 for the full MS spectrum followed by ten MS/MS spectra in the linear ion trap using 35 as threshold energy of for collision-induced dissociation.

For the targeted analysis of the GFP tryptic peptide LEYNFNSH[X]VYITADK, where X=glutamine, serine, phosphoserine or **2**, the theoretical masses were determined to 4 significant figures for all relevant charge states and the raw data searched with a m/z range of ±0.15 Da. The resulting extracted ion chromatograms were integrated and the area-under-the-curve (AUC) was used for relative quantitation.

The collected LC-MS/MS data were searched using the Mascot search engine (Matrix Science) against database containing Swiss-Prot and the GFP construct with all amino acids and post-translational modifications allowed at position 150. Precursor tolerance of 5 p.p.m. and a fragment ion mass tolerance of 0.8 Da were used for the search, allowing two missed enzyme cleavages and variable modifications for oxidized methionine, carbamidomethyl cysteine, pyroglutamic acid and phosphorylated serine, threonine and tyrosine. MS/MS data were subsequently validated using the Scaffold program (Proteome Software Inc.) and interrogated manually.

2.2.5.2.10 MEK1 Expression

HEK293/PSAT-KO cells were seeded and transfected as described above with combination of pPB_FLAG-SepRS_Myc-EF1a_V5-eRF1(E55D)_V5-PSPH_4xU6-tRNA$^{v1.0}_{CUA}$, Erk-GFP reporter and pPB_MEK1(218TAG)_4xU6-tRNA$^{v1.0}_{CUA}$ or pPB_MEK1(218TAG222Asp)_4xU6-tRNA$^{v1.0}_{CUA}$. The cells were incubated in DMEM supplemented with 10% FBS and 10 mM **2**, lysed and analysed by phos-tag-SDS-PAGE or SDS-PAGE and western blotting, as described above.

2.2.5.3 QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of western blots was carried out using ImageJ. Details of all statistical datasets are included in the corresponding figure legends.

Chapter 3 Genetic Code Expansion for Site-specific Live-cell Protein Labelling

Genetic code expansion has emerged as a unique tool to label proteins in live cells. This chapter provides an overview of genetic code expansion in the context of protein labelling and imaging and describes two major advances in protein labelling in mammalian cells.

First, we leverage recent developments in genetic code expansion in mammalian cells to densely label proteins with bright and photostable organic dyes and image them using super-resolution microscopy.

In the second section, we identify the readthrough of endogenous amber codons, resulting in aberrantly extended proteins bearing the non-canonical amino acid, as the primary source of non-specific labelling. We minimise this off-target labelling through optimisation of the transfection protocol for multiple distinct proteins. We then demonstrate that the resulting protocol allows live-cell imaging of low levels of proteins with signal-to-noise comparable to immunofluorescence.

In the last section, we demonstrate the efficiency of the new protocol for live-cell imaging, describing the dynamic behaviour of a recently discovered microprotein NoBody in live cells.

3.1 Introduction

Visualization of biomolecules by fluorescent microscopy has transformed our understanding of biological systems (Tsien, 1998). The use of fluorescence, in which the imaged object itself emits light after absorbing it, allows for better image quality as the emitted light can be spectrally separated due to the Stokes shift that results as vibrational relaxation of the excited state before the photon is emitted. When coupled with the right labelling technique fluorescence imaging provides molecular specificity to the otherwise very complex biological samples.

Fluorescent imaging has been useful across several orders of magnitude of both space and time: from morphogenesis of tissues and organs in entire organisms such as the fruit fly (McMahon et al., 2008), nematode (Liu et al., 2009), zebrafish (Keller et al., 2008) or mouse embryos (Keller, 2013; Trichas et al., 2012) over many days down to structural insights on the nanometre scale (with most recent super-resolution methods) and nanosecond time resolution (Elangovan et al., 2002; St-Pierre et al., 2014).

3.1.1 Super-resolution imaging

The advent of super-resolution microscopy has enabled researchers to observe biological processes occurring on a scale below the diffraction limit (Betzig et al., 2006; Klar et al., 2000; Rust et al., 2006). In aggregate the two main approaches single-molecule localization microscopy (which includes photoactivated localization microscopy (Betzig et al., 2006) and stochastic optical reconstruction microscopy (Rust et al., 2006)) and stimulated emission depletion microscopy (STED) (Klar et al., 2000) use photophysical properties of the fluorescent dyes to break the diffraction barrier, first described by Abbe in 1873 (Abbe, 1873). Patterned illumination approaches also allow improvement in resolution by expanding the optical transfer function of the microscope ((Gustafsson, 2005)), however, the resolution improvement is limited to a factor of 2 (in all 3 simensions).

Specifically, the methods rely on turning majority of the fluorescent molecules to an OFF state. STED achieves this by stimulated emission in a structured ‘donut’ pattern around the focal

point (Hell and Wichmann, 1994), which allows for saturable depletion outside of the focal point. The donut pattern is achieved by insertion of phase mask into the stimulated emission beam which itself is superimposed on the excitation beam. As a result fluorophores positioned anywhere outside of the very centre of the 'donut' beam are forced into ground state via the stimulated emission. Subsequently, the fluorescence from the subdiffraction area inside the donut is detected (Hell and Wichmann, 1994; Klar et al., 2000). The single-molecule localisation microscopy approaches illuminate the entire sample using a widefield microscope but use special conditions where only small subset of fluorophores emit at any given time (Huang et al., 2009). In PALM or fluorescence PALM (fPALM) (Hess et al., 2006) molecules are stochastically converted by low-power activating beam into active state, where they are imaged and immediately photobleached back into non-fluorescent state (Betzig et al., 2006). This is possible due to the discovery of several variants of green fluorescent protein (GFP) that can be activated or change colour by high-frequency (~ 405 nm) light (Ando et al., 2002; Patterson and Lippincott-Schwartz, 2002; Wiedenmann et al., 2004). In STORM, majority of the molecules are excited but subsequently transition into the dark triplet state in the presence of thiols (usually mercapthoethylamine or β-mercaptoethanol) (Dempsey et al., 2011).

The photophysical properties are critical for the performance of super-resolution methods. Main parameters are photon emission, the duty cycle (ratio of time spent in dark state vs. active state), photostability and number of switching cycles before bleaching (Dempsey et al., 2011). This demand has created an urgent need for efficient, specific, and non-perturbative methods to label molecules in living cells with bright and photostable fluorophores.

3.1.2 Fluorescent proteins

Historically, fluorescent labelling of organisms relied on antibody labelling (Coons, 1942), which requires sample fixation and, for imaging of intracellular proteins, membrane permeabilisation. Inherently, cell fixation does not provide information on cellular dynamics, which can be obtained through either direct imaging or alternative techniques such as fluorescence recovery after photobleaching (FRAP), fluorescence correlation spectroscopy (FCS), fluorescence cross-correlation spectroscopy (FCCS), single-particle tracking (SPT), or Förster resonance energy transfer (FRET) experiments. In addition, fixation of cells for

microscopy may lead to significant artifacts and distortions of cellular features (Schnell et al., 2012).

Since 1990s the green fluorescent protein (Chalfie et al., 1994; Prasher et al., 1992) and its spectral variants (Heim and Tsien, 1996) have ignited a revolution in live-cell imaging (Tsien, 1998), thanks to the ability to encode the fluorescent protein tag genetically as a fusion to the protein of interest. This development has allowed scientists to study protein localisation using fusion tags, use GFP as a reporter gene, in FRET exepriments, for metabolite indicators, pH indicators and many others (Zimmer, 2002). Despite many developments, the nature of the fluorophore formation in fluorescent proteins (FP) inherently limits their photophysical properties (Wang et al., 2014).

3.1.3 Organic fluorophores

Organic fluorophores provide an attractive alternative to fluorescent protein due to their virtually unconstrained chemical structure. Dye synthesis is a traditional field of organic chemistry and the majority of fluorescent dyes such as coumarins, fluoresceins, rhodamines phenoxazines and cyanines were synthesized in the late 19^{th} or early 20^{th} century (Lavis, 2017). To improve the dye properties various functionalities were introduced, such as sulfonation to increase solubility or structural motifs to rigidify the structure to increase photostability and brightness. The resulting lines of dyes, including the dyes known as ‘Alexa Fluor’ (Panchuk-Voloshina et al., 1999) and ‘CyDyes’ (Mujumdar et al., 1993), are still commonly used today and have been modified for various uses including staining of various organelles, or as ion indicators (Lavis and Raines, 2008).

The recent advances in super-resolution microscopy with its high demand on the ‘photon budget’ have spurred new developments in organic dyes. Most notably, silicon-substituted rhodamine (SiR) has emerged as a particularly useful, far-red dye for super-resolution imaging (Lukinavičius et al., 2013). Besides its excellent spectroscopic properties, SiR exists in equilibrium between open, zwitterionic and closed forms, which is nonfluorescent. The free dye in solution adopts the closed form, but shifts the equilibrium when bound to a polar protein surface. This renders the SiR a ‘fluorogenic’ dye and leads to minimal background originating

from an unbound dye, reducing the necessity for sample washes (Lukinavičius et al., 2013). SiR has since been engineered for, among other uses, labelling of cytoskeletal proteins (Lukinavičius et al., 2014) and DNA (Lukinavičius et al., 2015) and further developed to a further red-shifted version SiR700 (Lukinavičius et al., 2016). An alternative to SiR derivatives have been developed by the Lavis group, who have used Pd-catalyzeed cross-coupling chemistry to introduce azetidine ring into tetramethyl-rhodamine, significantly increasing its brightness and photostability (Grimm et al., 2015), an approach that could be further extended to other dyes spanning the visible spectrum (Grimm et al., 2015; Grimm et al., 2017).

3.1.4 Current protein labelling strategies

Currently, the approaches for specific protein labelling with organic dyes rely on genetically encoded tag expressed as part of the protein, which directs the fluorophore to the protein. The commonly used protein tags include SNAP-tag/CLIP-tag (Gautier et al., 2008), HaloTag (Los et al., 2008), FlAsH (Martin et al., 2005), and PRIME (Uttamapinant et al., 2010). All of these methods rely on attachment of protein or peptide appendages to the target molecules, which are known to perturb their function, localisation and dynamics. In addition, these methods are limited to the protein termini and, in some cases, flexible loop domains, which renders many important targets impossible to visualise in live cell context either without major perturbations or completely.

An alternative to protein fusion is to use a small molecule binder. In an example of this approach, Lukinavicius and colleagues fused the silicon-rhodamine dye to paclitaxel (Taxol) and phalloidin, known microtubule and actin binding molecules (Lukinavičius et al., 2014). The resulting molecules allowed them to study the cell cytoskeleton using super-resolution microscopy without any genetic modification. While simple to use, the small molecule binders must be membrane permeable and not actively exported from the cell interior, which is not trivial (Lukinavičius et al., 2014). Similarly, this approach is not easily transferrable to other protein targets for which a strong and specific small molecule binder may not be available. Lastly, the targeting moiety must not interfere with the target protein function.

3.1.5 Bioorthogonal chemistry and genetic code expansion

Genetic code expansion has recently emerged as a unique method to label proteins in live cells with organic fluorophores. This development was enabled by the advance in bio-orthogonal chemistry, a group of chemical reactions that:

(1) are selective under physiological conditions and with respect to the cellular environment;
(2) produce stable covalent linkage that is not metabolically or thermodynamically broken and;
(3) produces no toxic byproducts and require no toxic catalysts (Lang and Chin, 2014).

In the live-cell labelling context, these reactions should also have high second-order rate constant k_2 (labelling reactions normally follow second-order kinetics) such that quantitative labelling can be achieved on a reasonable timescale without necessity to provide an excess of labelling reagent. Among the most widely used reactions for cellular labelling are azide-alkyne reactions ([3 + 2] cycloadditions) (Himo et al., 2005). To proceed at physiological conditions, the cycloaddition requires the use of copper(I) salts (Tornøe et al., 2002; Wang et al., 2003), which can be toxic to cells (Wolbers et al., 2006). An attractive alternative are the strained alkene/alkyne-tetrazine reactions ([4 + 2] Diels-Alder cycloadditions) (Boger, 1986), which proceed at much higher rate than strain-promoted click reactions.

As reviewed in section 1.6, genetic code expansion uses a new, orthogonal aminoacyl-tRNA synthetase/tRNA pair to incorporate non-canonical amino acid into proteins, generally in response to the amber stop codon. Pioneered by the Schultz group, several amino acids bearing bio-orthogonal handles, such as azides, alkynes, alkenes or ketones have been incorporated (Chin et al., 2003; Chin et al., 2002b; Deiters et al., 2003; Mehl et al., 2003; Nguyen et al., 2009b; Wang et al., 2003; Zhang et al., 2002). More recently, strained alkenes, such as norbornenes and trans-cyclooctenes (Lang et al., 2012a) and bicycle[6.1.0]nonynes (Lang et al., 2012b) have been incorporated and used to site-specifically label proteins in *E. coli* and mammalian cells with tetrazine-dye conjugates. In many instances the use of tetrazine conjugated dyes exhibit weak fluorescence when in solution but become strongly fluorescent upon attachment (Lang et al., 2012b). This fluorogenic behaviour further helps to increase the

specificity of the labelling approach by supressing the background and limiting the need for sample washes.

3.2 Genetic code expansion enables live-cell and super-resolution imaging of site-specifically labeled cellular proteins

3.2.1 Introduction

The rise of super-resolution microscopy led to a renewed interest in protein labelling technologies. As introduced in chapter 3.1, super-resolution microscopy relies on a bright and photostable fluorophores that provide large 'photon budget', allowing imaging with high spatial and temporal resolution whilst maintaining low phototoxicity (Liu et al., 2015). Due to the very high spatial resolution of the imaging methods, the size of the fluorescent probe itself can play important role. The current experimental limits of approx. ~10 nm in X and Y (Strauss et al., 2018) are on the same range as the size of fluorescence proteins (~ 4.2 x 2.4 nm) and significantly below the size of commonly used antibodies (10 – 15 nm for single antibody and up to 30 nm when using secondary antibodies). Thereby, labelling proteins directly with organic fluorophores, which are small, bright and very photostable, offers a significant advantage for super-resolution imaging.

Super-resolution microscopy may be used to determine the sub molecular structure of protein assemblies through particle averaging (Szymborska et al., 2013). To achieve this, multiple sites on the target protein or protein assembly need to be labelled and thereby specific antibodies against various epitopes need to be raised. The genetic code expansion, which allows site-specific labelling, is ideally suited to this approach, extending its applicability to virtually any proteins or protein assemblies.

Simultaneously, the high image resolution requires a dense labelling of the ultrastructures to achieve sufficient spatial resolution (Nieuwenhuizen et al., 2013). Previously, genetic code expansion in mammalian cells suffered from poor yields, making high level of protein expression and labelling difficult. Recently, Schmied et al. have developed a new mammalian expression system, which provided high expression thanks to high level of the amber suppressor $tRNA^{Pyl}_{CUA}$. This system expressed 129% of full length GFP containing a single non-canonical amino acid (Nε-Boc-L-lysine) when compared to a wild-type GFP expressed from cytomegalovirus (CMV) promoter (Schmied et al., 2014)

3.2.2 Super-resolution imaging of mammalian cytoskeleton

Using the mammalian expression system developed by Schmied and colleagues, Uttamapinant et al. were able to densely label and cytoskeletal proteins actin and vimentin as well as the epidermal growth factor (EGFR) and image them using multiple super-resolution techniques (Uttamapinant et al., 2015). Both actin and vimentin were labelled by incorporating the bicyclo[6.1.0]nonyne-lysine (BCNK) via the BCNKRS/tRNA$^{Pyl}_{CUA}$ pair (Lang et al., 2012b). Once expressed, the proteins were labelled in live cells with 4-(1.2.4.5-tetrazin-3-yl)benzylamino derivative of silicon rhodamine (SiR-tetrazine) (Lukinavičius et al., 2013), via the inverse-electron-demand Diels-Alder cycloaddition (**Figure 3.1**).

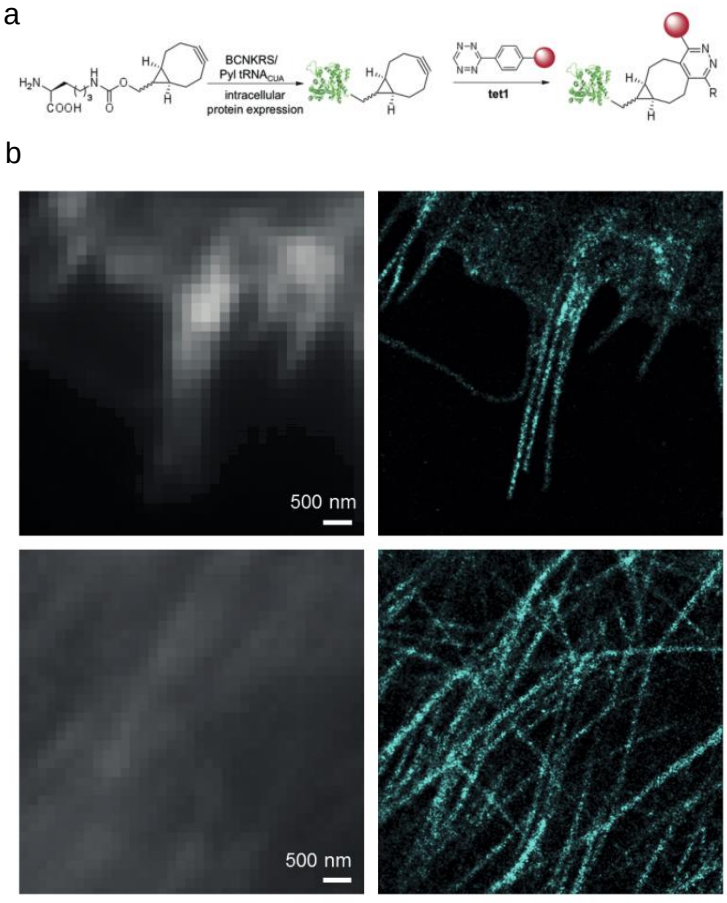

Figure 3.1. Super-resolution imaging of mammalian cytoskeleton via genetic code expansion. (a) Protein labelling via incorporation of BCNK using the BCNKRS/$tRNA^{Pyl}_{CUA}$ pair (Lang et al., 2012) and subsequent derivatization with tetrazine-dye conjugate. (b) Representative image of actin(K118BCNK) using widefield (left panels) and STORM microscopy (right panels). Adapted from (Uttamapinant et al., 2015).

3.2.3 $tRNA^{Pyl}_{CUA}$-dependent labelling background

When imaging the cytoskeletal protein vimentin, we observed an occasional fluorescent signal in the nucleus of some cells. Because vimentin is an exclusively cytosolic protein, we hypothesized that the signal may be originating from aminoacylated $tRNA^{Pyl}_{CUA}$ trapped in the nucleus where it is not available for translation. To verify this, we carried out fluorescence *in situ* hybridisation (FISH) experiments, to see whether the background signal co-localises with the $tRNA^{Pyl}_{CUA}$. We transfected HEK293 cells with plasmids containing the BCNKRS, $tRNA^{Pyl}_{CUA}$ and Myc-vimentin(N116TAG) and incubated them in presence of 1 mM BCNK for 48 hours. The cells were washed and labelled with 400 nM carboxyfluorescein diacetate (CFDA) tetrazine conjugate and fixed. The cells were then stained with Alexa Fluor 647 labelled oligonucleotide targeting the $tRNA^{Pyl}_{CUA}$ and anti-Myc antibody to verify the localization of vimentin. We observed that indeed the $tRNA^{Pyl}_{CUA}$ signal co-localised with the nuclear CFDA signal but not with the anti-Myc signal, confirming that the source of the non-specific signal is the aminoacylated $tRNA^{Pyl}_{CUA}$ (**Figure 3.2**). While this problem may be remedied by longer washout periods, allowing the aminoacylated tRNA to be turned over, this situation posses a challenge for proteins with high turnover or protein synthesized at a low level.

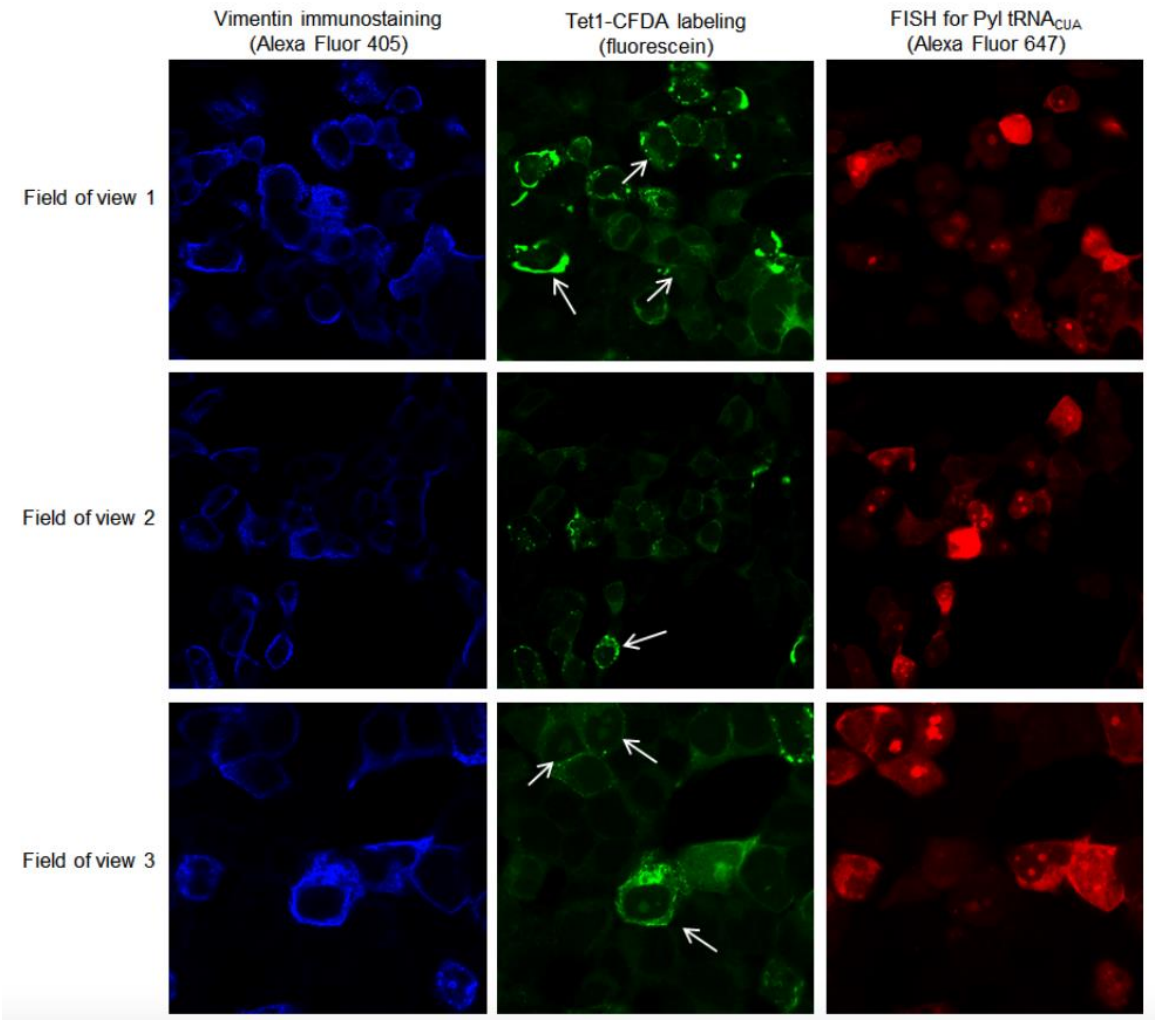

Figure 3.2. $tRNA^{Pyl}_{CUA}$-dependent labelling background. HEK293 cells transfected with plasmids bearing BCNKRS, $tRNA^{Pyl}_{CUA}$ and Myc-vimentin(N116TAG) were cultured in presence of 1 mM BCNK and labelled with tetrazine-CFDA (centre panels), oligonucleotide targeting the $tRNA^{Pyl}_{CUA}$ (right panels) and anti-Myc antibody, targeting the product of Myc-vimentin(N116TAG) gene (left panels). White arrows highlig1ht the nuclear background co-localizing with $tRNA^{Pyl}_{CUA}$ signal but not anti-Myc staining. (Figure taken from (Uttamapinant et al., 2015))

3.3 Optimizing live-cell labelling signal-to-noise ratio

3.3.1 Optimization of labelling in mammalian cells

In previous experiments we (Uttamapinant et al., 2015) and others (Peng and Hang, 2016) have observed low level of background labelling of the mammalian proteome when mammalian cells were transfected with one of the evolved variants of pyrrolysyl-tRNA synthetase (PylRS)/pyrrolysyl-tRNA$_{CUA}$ (tRNA$^{Pyl}_{CUA}$) pair from *Methanosarcina* species (MbBCNKRS (Lang et al., 2012b) or *Mm*PylRS-AF mutant (Yanagisawa et al., 2008)) in combination with an amber mutant of a protein of interest.

Because this non-specific labelling is dependent on the presence of the non-canonical amino acid with bioorthogonal reactive group and manifested as several discrete bands on fluorescent scans of SDS-PAGE gels, we hypothesised that the non-specific background originates from aberrant readthrough of endogenous amber codons rather than non-specific attachment of dye molecules to cell proteome. While aberrantly extended proteins are degraded by the cell over time (Arribere et al., 2016), this background can lead to decrease in signal-to-noise ratio for in vivo imaging applications, especially on shorter timescales.

Two recent advances in protein labelling using non-canonical amino acids focus on decreasing the labelling of the charged *Mm*Pyl-tRNA$^{Pyl}_{CUA}$ or the *Mm*PylRS/tRNA$^{Pyl}_{CUA}$ pair complex. The first approach relies on simple decrease of the tRNA$^{Pyl}_{CUA}$ expression (Aloush et al., 2018), the latter approach relies on observation that the PylRS from *Methanosarcina mazei* (*Mm*) locates to the nucleus when expressed in mammalian cells and consequent trapping of the charged tRNA$^{Pyl}_{CUA}$, which is labelled by the tetrazine-dye conjugate (Nikić et al., 2016). The solution presented was the addition of nuclear export sequence (NES) to the coding sequence of the PylRS, that resulted in up to 15-fold increase in amber suppression efficiency along with marked decrease of labelling background.

We tested this approach using the amino acid BCNK, which permits significantly more efficient labelling of the expressed proteins and comparable expression rate (Peng and Hang, 2016), while preventing chemical elimination of the functional group of the amino acid (Ge et

al., 2016). Unfortunately, under the condition we tested (5 and 50 μM BCNK) the several different N and C-terminal fusions of NES did not show improved amber suppression efficiency compared to that of the wild type *Mm*PylRS-AF (**Figure 3.3** and **3.4**).

Next we checked for the presence of nuclear aggregation using the *Mm*PylRS-AF mutant. We overexpressed FLAG tagged versions of the *Mm*PylRS-AF and NES-*Mm*PylRS-AF in combination with $tRNA^{Pyl}_{CUA}$ in HEK293 cells. The cells were grown in the presence of 50 μM BCNK and labelled using SiR-tetrazine. The cells were subsequently fixed, permeabilised, immunostained with anti-FLAG antibody and imaged using confocal microscope. Similarly to our previous observations (**Figure 3.2**) and previous studies (Nikić et al., 2016) we detected discernable nuclear aggregation of the *Mm*PylRS-AF in the form of puncta (**Figure 3.3**). Some of the nuclear aggregates co-localised with labelling in the SiR channel, suggesting that nuclear-trapped BCNK-$tRNA^{Pyl}_{CUA}$ can be labelled by the tetrazine-dye conjugate (**Figure 3.3**). When using the NES-*Mm*PylRS-AF we did not observe any nuclear aggregates of NES-*Mm*PylRS-AF supporting this hypothesis, however, the major fraction of the non-specific labelling persisted (**Figure 3.3**). This observation is consistent with the hypothesis that majority of the labelling originates from the aberrantly extended endogenous proteins. We note that we observed nuclear puncta only when we overexpressed the *Mm*PylRS-AF (**Figure 3.5**) and under the conditions of our labelling experiments we did not see significant aggregation and labelling of $tRNA^{Pyl}_{CUA}$ in the cell nuclei (**Figure 2, S3, S6, S7**).

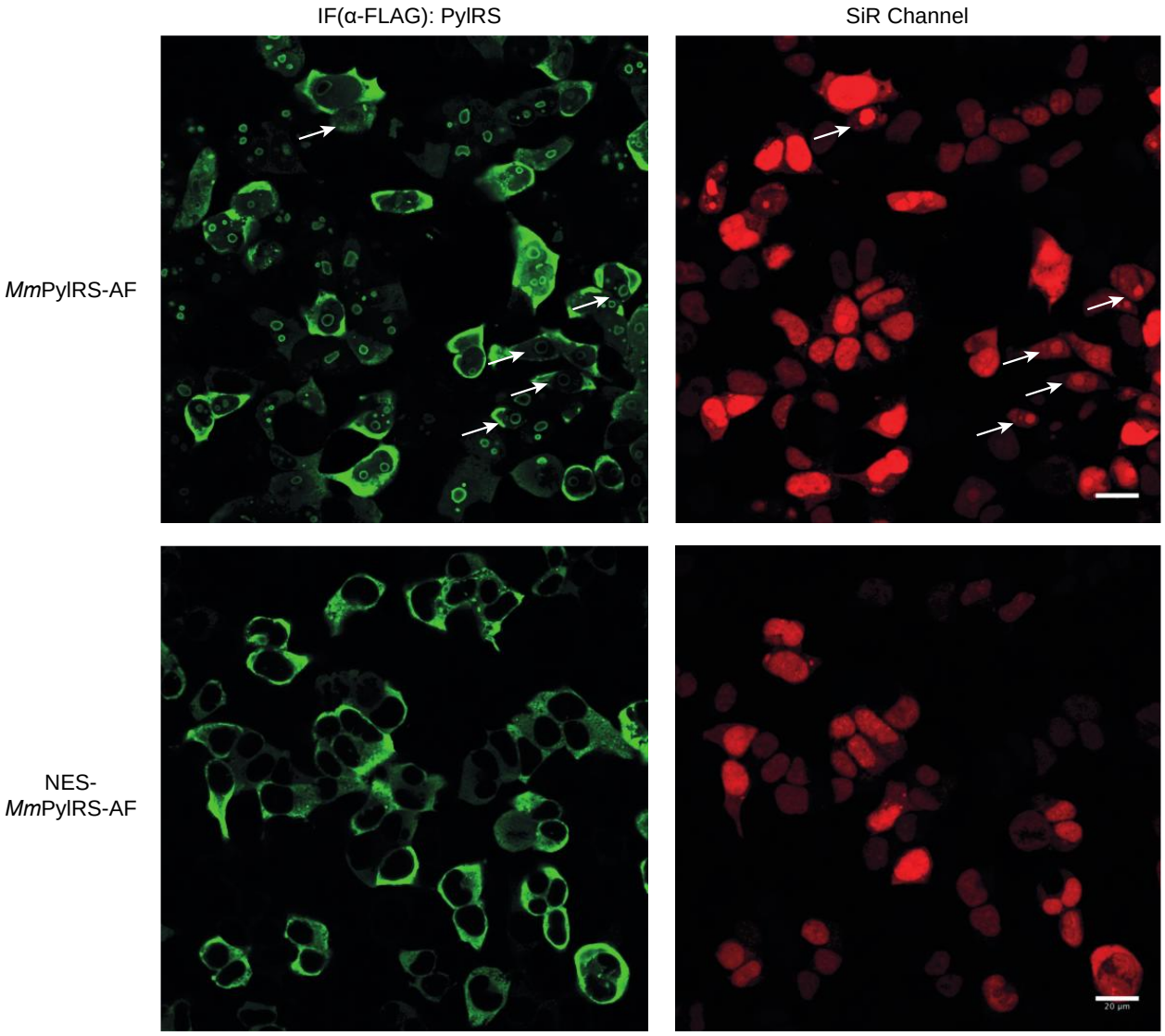

Figure 3.3. Effect of nuclear export of MmPylRS-AF on cellular localization and nuclear labelling background. HEK293 cells transfected with *Mm*PylRS-AF (top panels) and NES-*Mm*PylRS-AF (FLAG-NS2-*Mm*PylRS-AF, see Figure S2) (bottom panles).The cells were cultured in presence of BCNK for 24 hours, washed and labelled with SiR-tetrazine. Nuclear aggregates observed when *Mm*PylRS is overexpressed are highlighted with white arrows. Scale bar represents 20 µm.

Similarly, we checked for the presence of nuclear aggregation and observed that under the conditions used in our study, we did not see significant aggregation of ncAA labelling in the cell nuclei but rather a diffuse signal present throughout the cell (**Figures 3.5, 3.8, 3.9 and**

3.10). When we overexpressed the *Mm*PylRS-AF we did observe occasional aggregation in the cell nucleus. Under these conditions, the labelling on the $tRNA^{Pyl}_{CUA}$ may make a appreciable contribution to the non-specific background, and may be prevented by the use of NES.

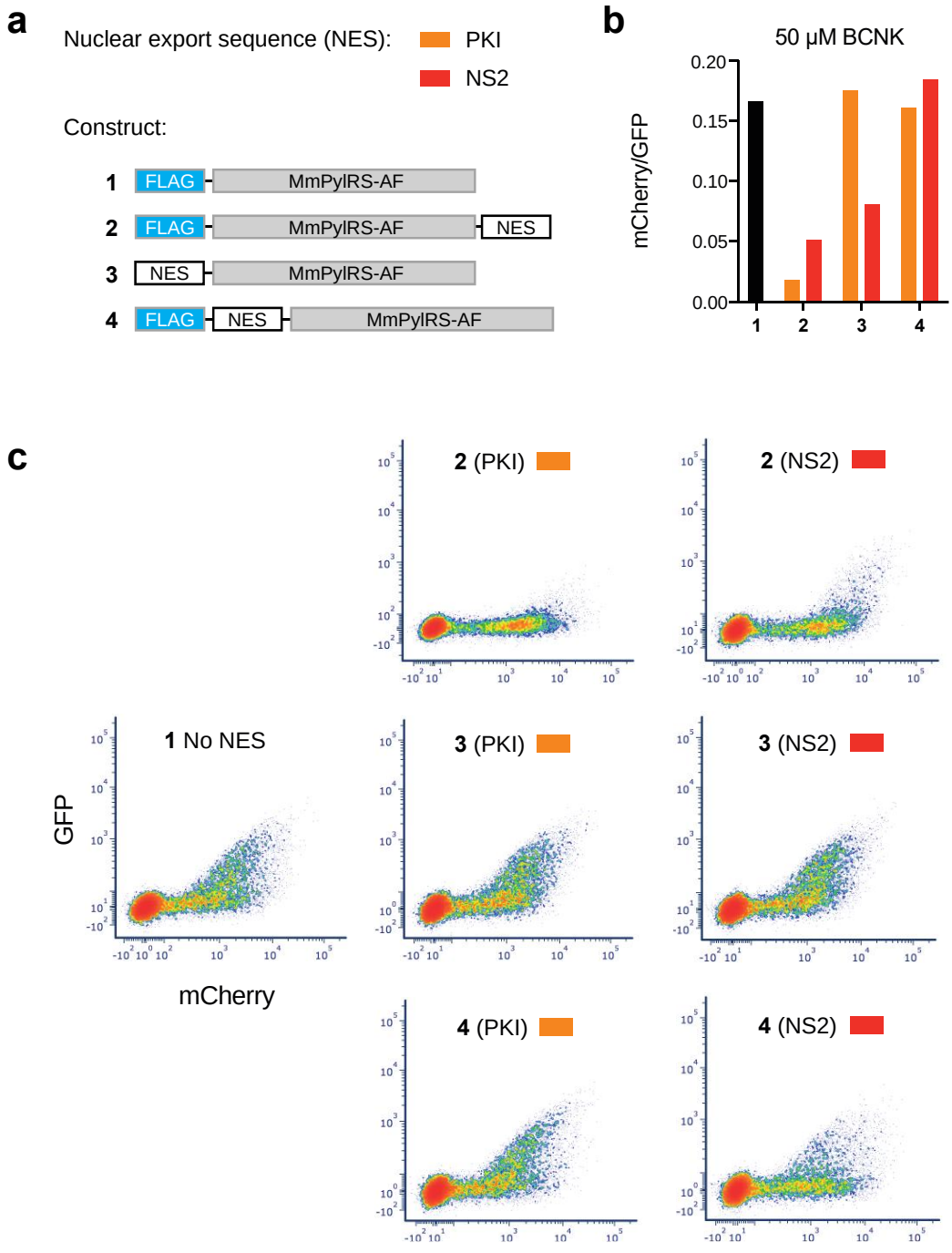

Figure 3.4. Effect of nuclear export of MmPylRS-AF on amber suppression at 50 µM BCNK. (**a**) Nuclear export sequences (NES) tested: PKI = LALKLAGLDIG and NS2 = MTKKFGTLTI. (**b**) Amber suppression efficiency measured by flow cytometry and expressed as a ratio of mCherry and GFP fluorescence resulting from co-transfection of mCherry-TAG-GFP reporter alongisde the PylRS-AF/tRNA$^{Pyl}_{CUA}$ transgenes. (**c**) Raw data used to calculate amber suppression efficiency in panel b.

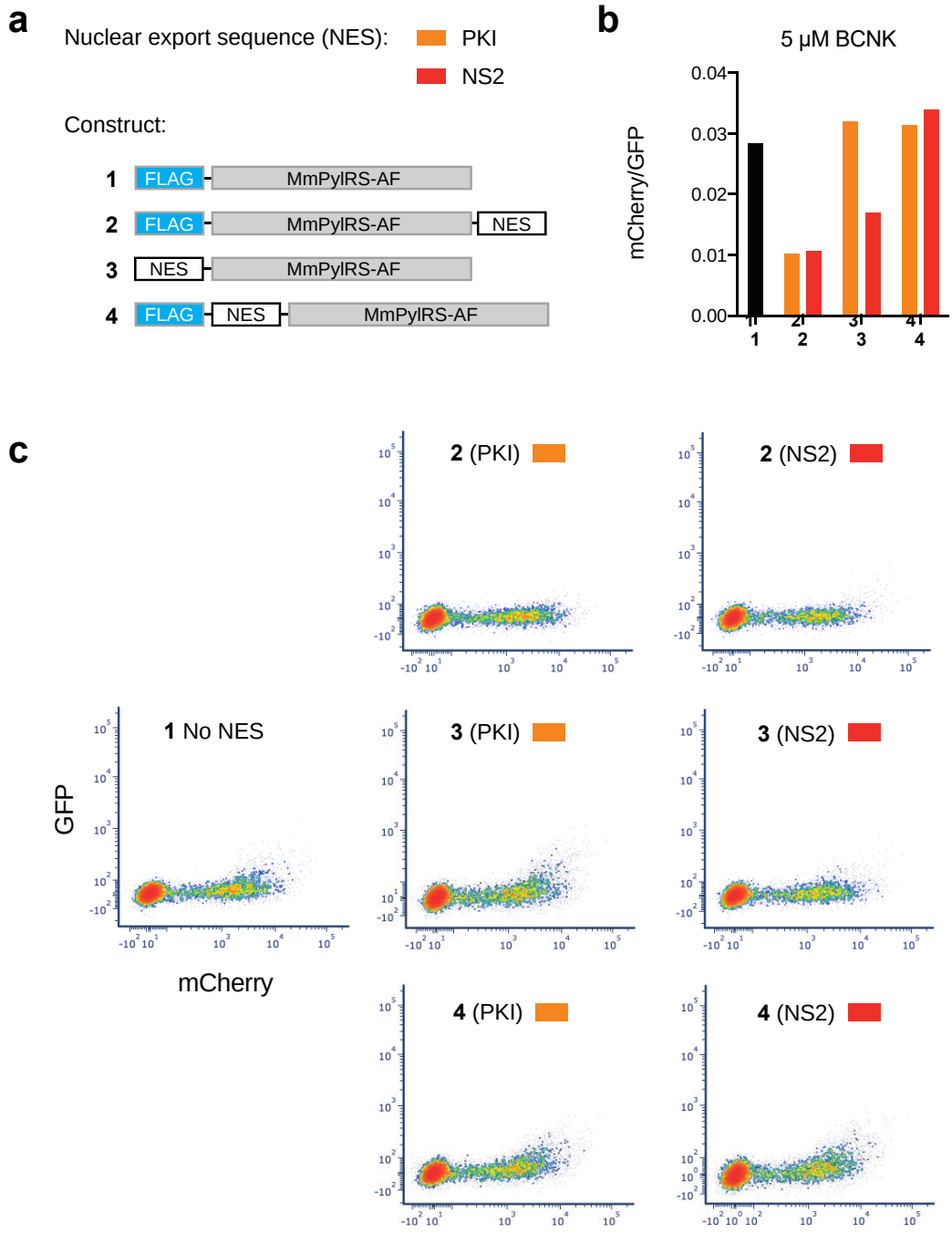

Figure 3.5. Effect of nuclear export of MmPylRS-AF on amber suppression at 5 μM BCNK. (**a**) Nuclear export sequences (NES) tested: PKI = LALKLAGLDIG and NS2 = MTKKFGTLTI. (**b**) Amber suppression efficiency measured by flow cytometry and expressed as a ratio of mCherry and GFP fluorescence resulting from co-transfection of mCherry-TAG-GFP reporter alongisde the PylRS-AF/ $tRNA^{Pyl}_{CUA}$ transgenes. (**c**) Raw data used to calculate amber suppression efficiency in panel **b**.

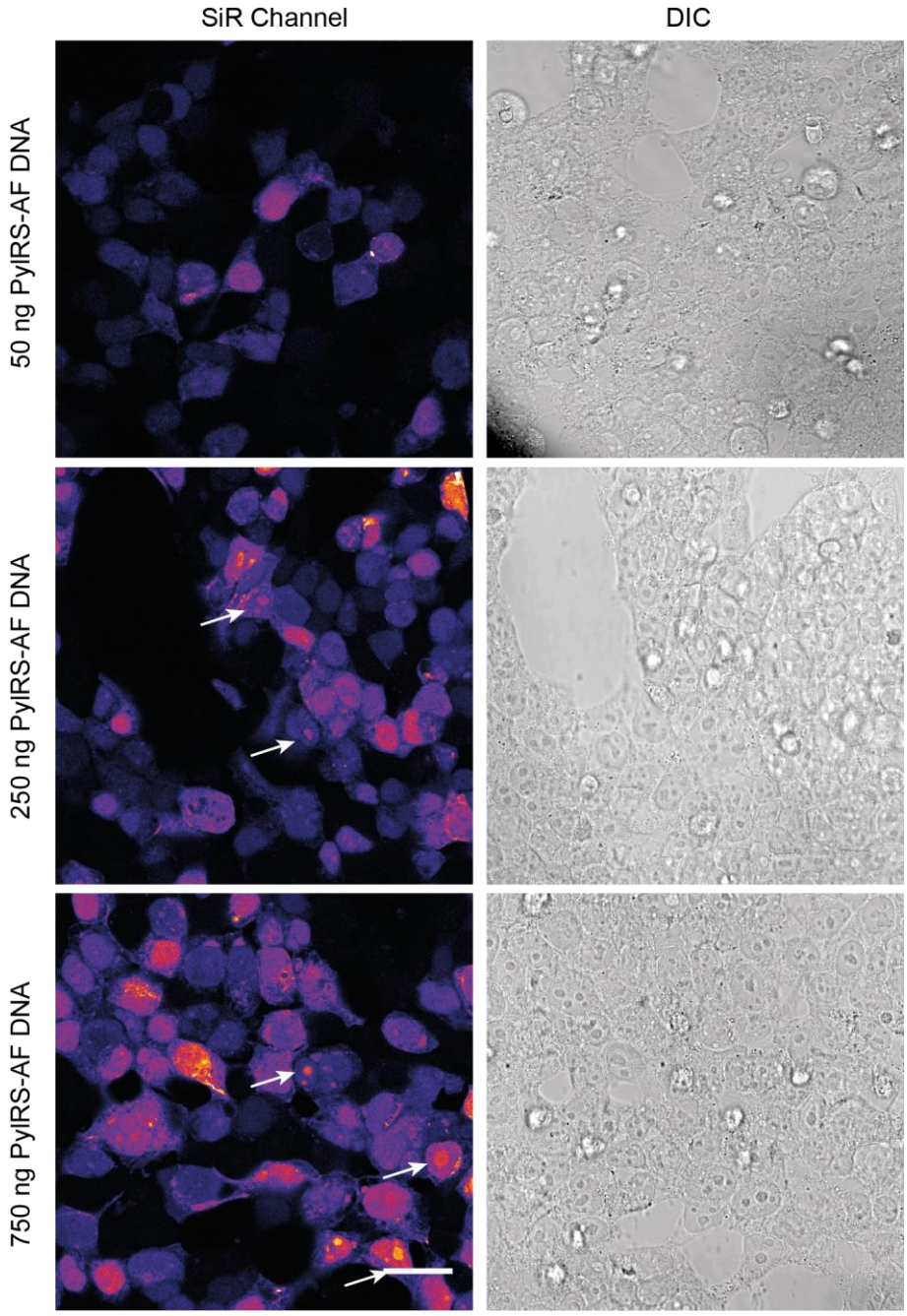

Figure 3.6. Nuclear background caused by aggregation of PylRS-AF in nucleus. HEK293 cells were transfected with 50 ng of vimentin(N116TAG) and variable amount of PylRS-AF plasmid, controlling for total amount of $tRNA^{Pyl}_{CUA}$ gene transfected and incubated in presence of 50 µM BCNK. Sub-nuclear puncta originating from PylRS-AF aggregation (white arrows) are apparent only in few cells and at high level of expression, while most of the background is diffuse throughout the nucleus. Scalebar 30 µm.

To improve the signal-to-noise ratio (SNR) we therefore sought to increase the specific signal, i.e. labelling of the protein of interest, while minimizing non-specific labelling of aberrantly extended proteins expressed by readthrough of endogenous amber codons. We hypothesized that increasing the level of mRNA of the target protein would lead to increase in specific signal as more of the charged $tRNA^{Pyl}_{CUA}$ is will be directed to ribosomes translating the transgene mRNA compared to readthrough of the endogenous amber codons. To test this hypothesis, we transfected HEK293 cells with increasing amount of sfGFP(150TAG) plasmid, maintaining constant amount of the *Mm*PylRS-AF and $tRNA^{Pyl}_{CUA}$ transgenes, controlling for the total amount of transfected DNA. After overnight expression in the presence of bicycle[6.1.0]non-4-yn-9-ylmethanol lysine (BCNK), the cells were washed and labelled for 20 minutes *in vivo* with 4-(1.2.4.5-tetrazin-3-yl)benzylamino derivative of silicon rhodamine (SiR-tetrazine). After additional washing step to remove the free dye, the cells were lysed and analysed by SDS-PAGE. Fluorescent scans of the gels (**Figure 3.7**) showed a dose-dependent increase of specific fluorescent labelling, corresponding to increase in expression of sfGFP(150BCNK), further confirmed by α-GFP western blot (bottom panel). Importantly, the increase in specific labelling was concurrent with decrease in background labelling of the proteome, consistent with the hypothesis that bigger fraction of the aminoacylated $tRNA^{Pyl}_{CUA}$ gets directed to translation of the sfGFP(150TAG) mRNA compared to the endogenous amber codons. Quantification of the signal confirmed 3-fold increase in SNR as a result of increased amount of sfGFP(150TAG) plasmid transfected (**Figure 3.7b** and **3.8**).

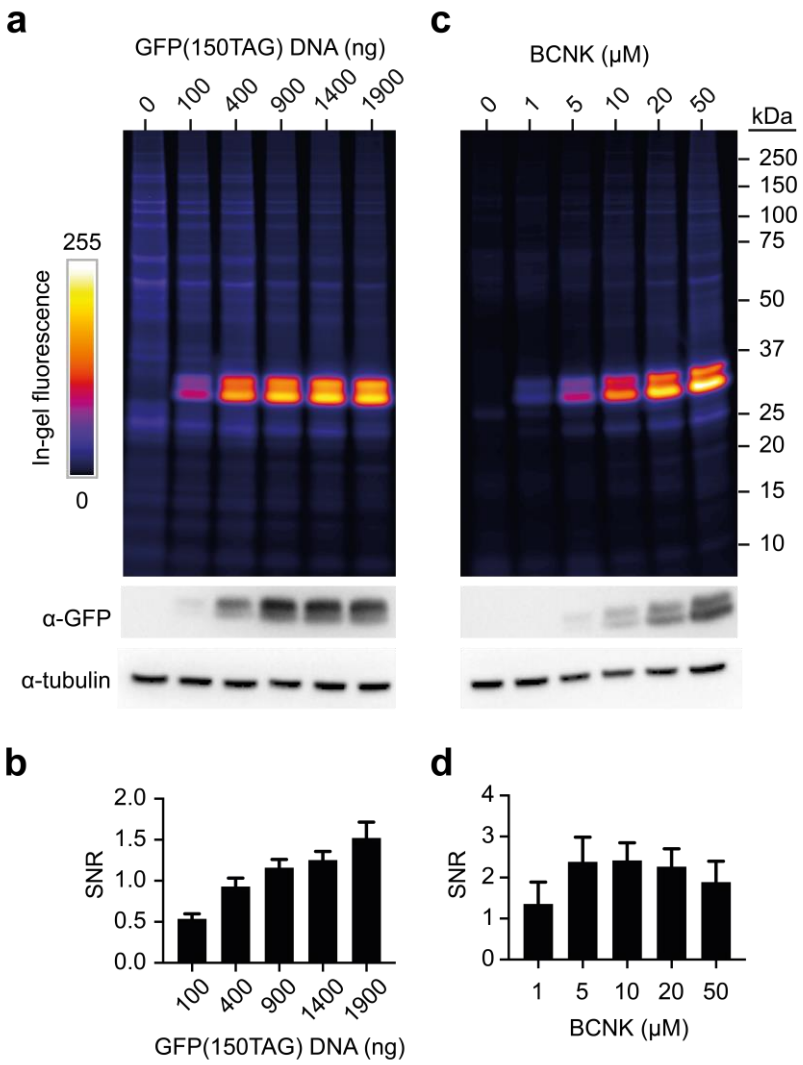

Figure 3.7. Optimization of protein labelling via genetic code expansion. **(a)** Fluorescent scan of SDS-PAGE gel of lysates of HEK293 cells transfected with increasing amount of sfGFP(150TAG) plasmid, incubated overnight in presence of 50 uM BCNK and labelled in vivo with SiR-tetrazine (top). α-GFP and α-tubulin western blot of the SDS-PAGE gel gel (bottom). The fluorescence contrast was enhanced through use of heat colormap (shown left) (see **Figure 3.8** for grayscale image fluorescent traces). **(b)** Quantification of signal-to-noise ratio based on fluorescence intensity of the GFP150BCNK band and the rest of the proteome. Data represent mean +/- SD from three biological replicates. **(c)** Fluorescent scan of SDS-PAGE gel of lysates of HEK293 cells transfected with 1900 ng of sfGFP(150TAG) plasmid, incubated overnight with varying amount of BCNK in the culture media and labelled in vivo with SiR-tetrazine (top). α-GFP western blot of the SDS-PAGE gel gel (bottom). The fluorescence contrast was enhanced through use of heat colormap (see **Figure 3.8** for grayscale image fluorescent traces). **(d)** Quantification of signal-to-noise ratio based on fluorescence intensity of the GFP150BCNK band and the rest of the proteome. Data represent mean +/- SD from three biological replicates.

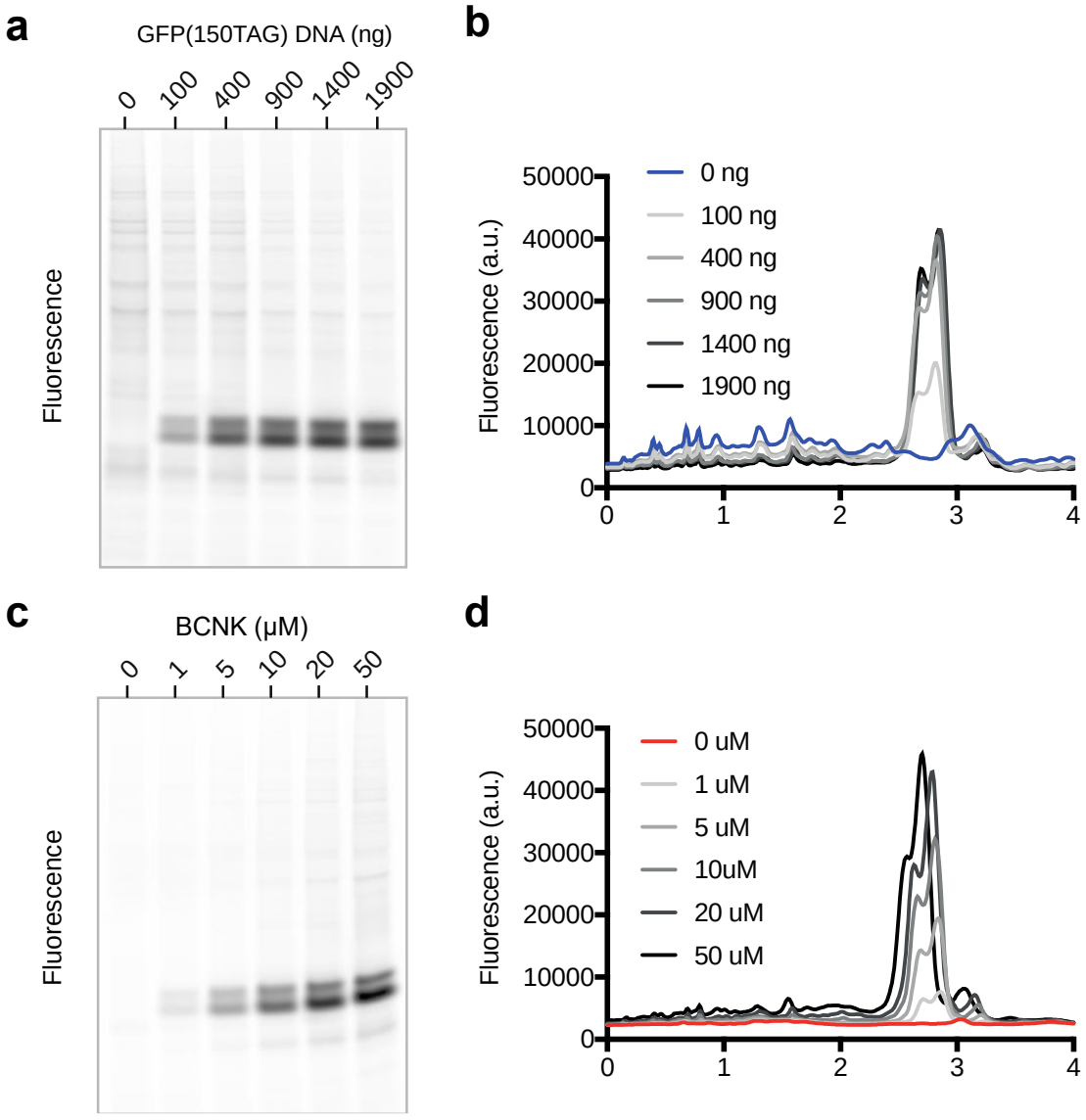

Figure 3.8. Grayscale versions of in-gel fluorescence scans and coresponding fluorescent traces. **(a) Figure 3.7a** in gray scale. **(b)** Fluorescent intensity traces of lanes in panel a. **(c) Figure 3.7c** in gray scale. **(d)** Fluorescence intensity traces of lanes in panel c.

Increasing the amount of transfected exogenous DNA may result in overexpression of the target protein. We hypothesized that decreasing the BCNK concentration will lead to decrease in both the expression and labelling of the protein of interest as well as aberrantly extended proteins with endogenous amber codons, resulting in constant SNR. We tested this hypothesis by transfecting HEK293 cells with sfGFP(150TAG) DNA and plasmid containing the *Mm*PylRS-AF and $tRNA^{Pyl}_{CUA}$ transgenes, varying the concentration of BCNK. As expected, both expression (**Figure 3.7c bottom**) and labelling (**Figure 3.7c top**) of sfGFP150BCNK was strongly dependent on the amount of supplied BCNK. Similarly, the labelling of the proteome decreased with the decreasing concentration of BCNK, consistent with lower readthrough of amber codons by the aminoacylated $tRNA^{Pyl}_{CUA}$. Consistent with our hypothesis, we observed a relatively constant SNR across the various concentrations of BCNK, with the exception of 1 µM BCNK, where other source of non-specific labelling may come into play (**Figure 3.7c top**). Importantly, this data suggests that lowering the concentration of BCNK can precisely regulate the level of target protein expression without compromising the SNR of the target labelling.

3.3.2 Verification for *in situ* labelling in mammalian cells

To confirm these results for *in situ* imaging, we took advantage of our observation that after transfecting cells with only *Mm*PylRS-AF and $tRNA^{Pyl}_{CUA}$ most of the fluorescent labelling resulting from aberrantly extended proteins in cells was located in the cell nucleus (**Figure 3.6**). We reasoned that transfecting and labelling an exclusively cytosolic protein would allow us to separate true signal located in the cytosol and the non-specific labelling in the nucleus and thereby quantitatively assess the SNR for *in situ* imaging. To this aim, we used previously reported Myc-vimentin(N116TAG) construct (Uttamapinant et al., 2015). We transfected HEK293 cells with increasing amount of Myc-vimentin(N116TAG) gene, keeping the amount of PylRS and $tRNA^{Pyl}_{CUA}$ genes as well as the total amount of transfected DNA constant. The cells were labelled with SiR-tetrazine, fixed and stained with α-Myc antibody, immunostaining only the exogenously expressed Myc-vimentin(N116BCNK) protein. We imaged the cells using confocal microscope, manually segmented the cells into nuclear and cytosolic parts and quantified the mean fluorescent signal originating from the two cellular compartments (**Figure 3.9a, 3.9b, 3.10a** and **3.11a**). For immunofluorescence, the SNR, calculated as a ratio of

cytosolic to nuclear signal, increased approximately 2-fold with increasing amount of DNA transfected, as expected from the increase in amount of Myc-vimentin(N116BCNK) expressed. For BCNK labelling, however, the SNR increased more than 8-fold between lowest and highest amount of Myc-vimentin(N116TAG) transfected (**Figure 3.9b**), with SNR comparable to immunostaining for transfection of 950 ng of Myc-vimentin(N116TAG) plasmid. We confirmed the imaging assay using SDS-PAGE analysis (**Figure 3.12**) with results in exact agreement with our *in situ* imaging assay.

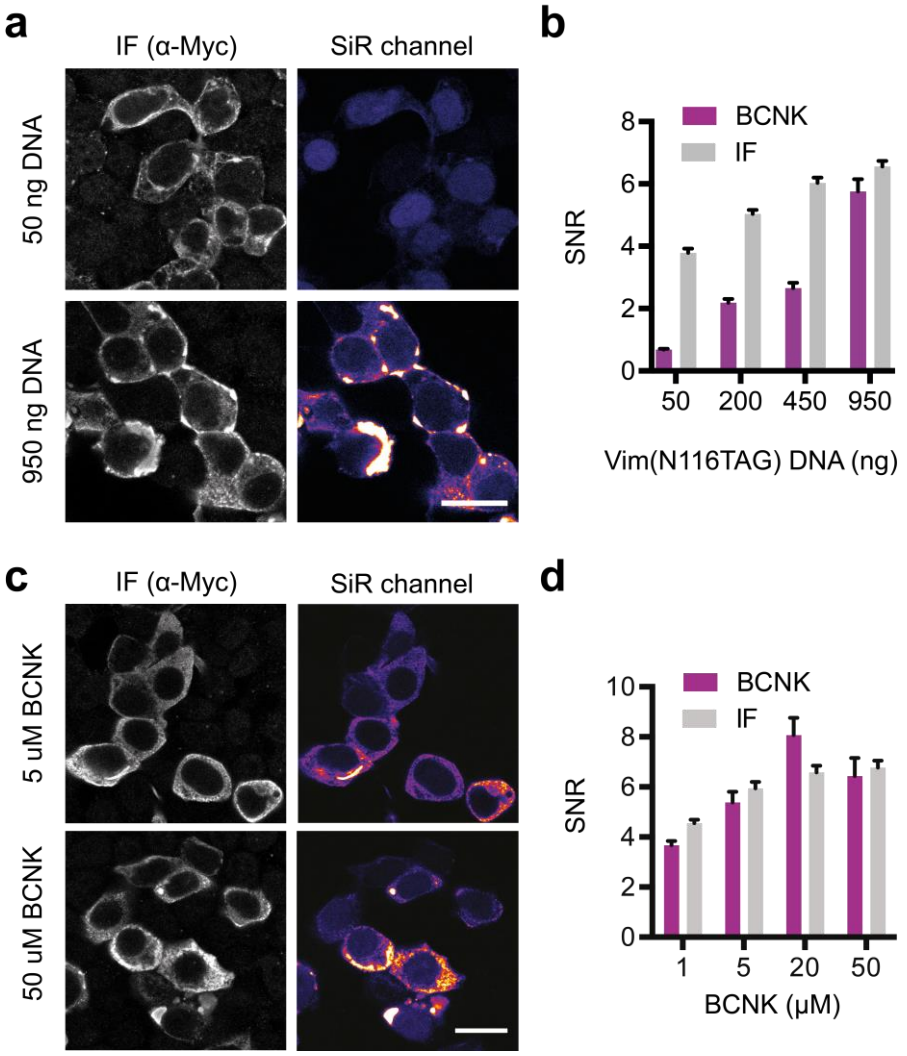

Figure 3.9. *In situ* quantification of signal-to-noise ratio for protein labelling via genetic code expansion. (a) Representative images of HEK293 cells expressing Myc-vimentin(116BCNK) labelled using α-Myc antibody (left) and SiR-tetrazine (right) after transfection of varying amount of the Myc-vimentin(116TAG) plasmid (for full scale images see **Figure 3.10**) (b) Signal-to-noise (SNR) calculated as ratio of mean cytosolic and nuclear signal (mean values ± SEM, n > 100 cells per condition). (c) Representative images of HEK293 cells expressing Myc-vimentin(116BCNK) labelled using anti-Myc antibody (left) and SiR-tetrazine (right) after incubation in presence of varying amount BCNK (for full scale images see **Figure 3.11**) (d) Signal-to-noise (SNR) calculated as ratio of mean cytosolic and nuclear signal (mean values ± SEM, n > 100 cells per condition).

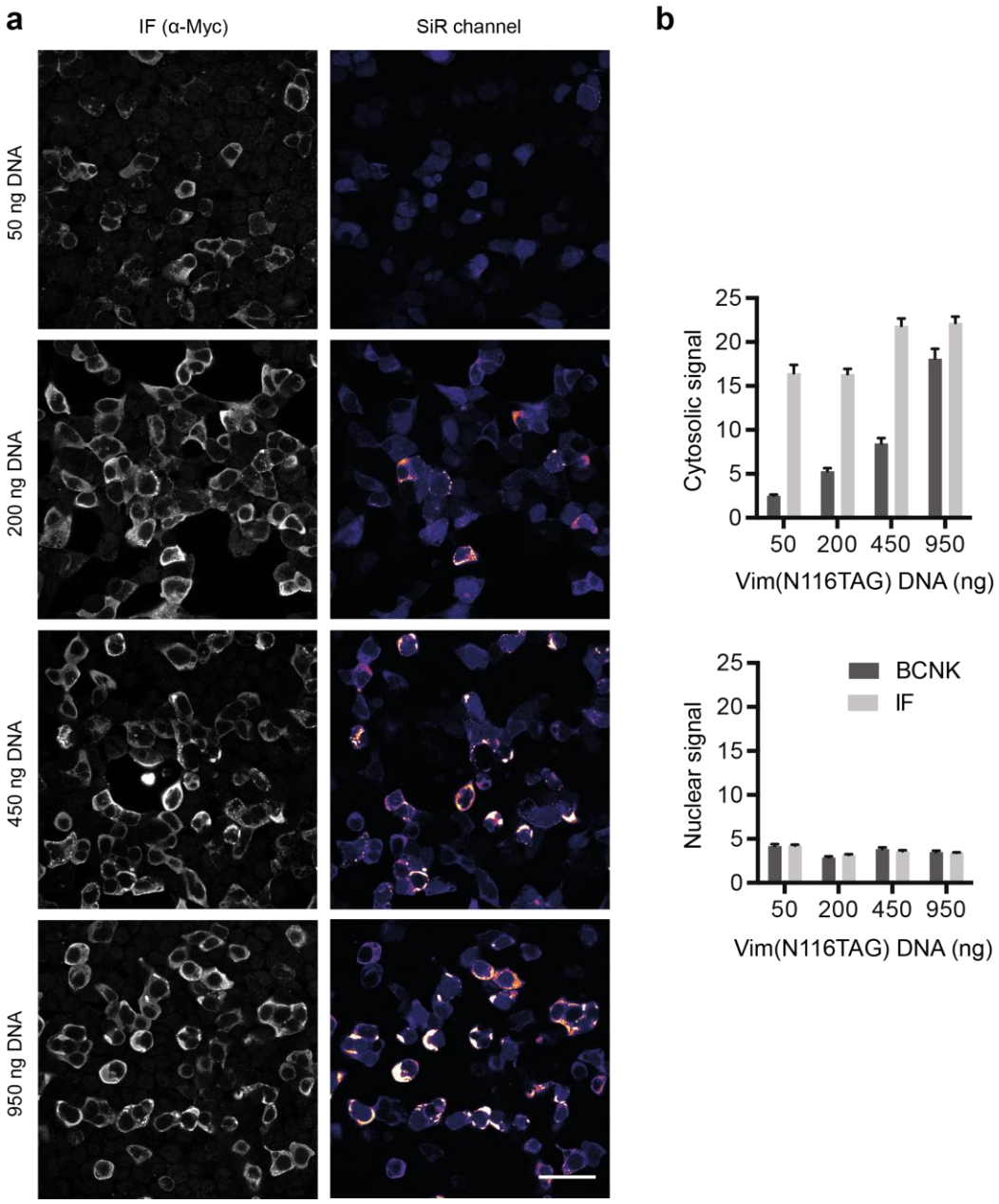

Figure 3.10. Reresentative full-size images shown in **Figure 3.9a** and used to generate data in **Figure 3.9b**. (a) HEK293 cells were transfected with progressively increasing amount of vimentin(N116TAG) DNA plasmid, keeping the amount of PylRS-AF and $tRNA^{Pyl}_{CUA}$ genes constant. The vimentin(116BCNK) was labellied in vivo by SiR-tetrazine (SiR channel) and post fixation and permeabilisation by anti-Myc antibody (IF (α-Myc)). Scalebar 50 μm. (b) Cells were manually segmented into nuclear and cytoplasmic regions and the signal in both BCNK and IF channel was quantified. Values shown are mean ± SEM. Each bar represents over 100 cells analysed.

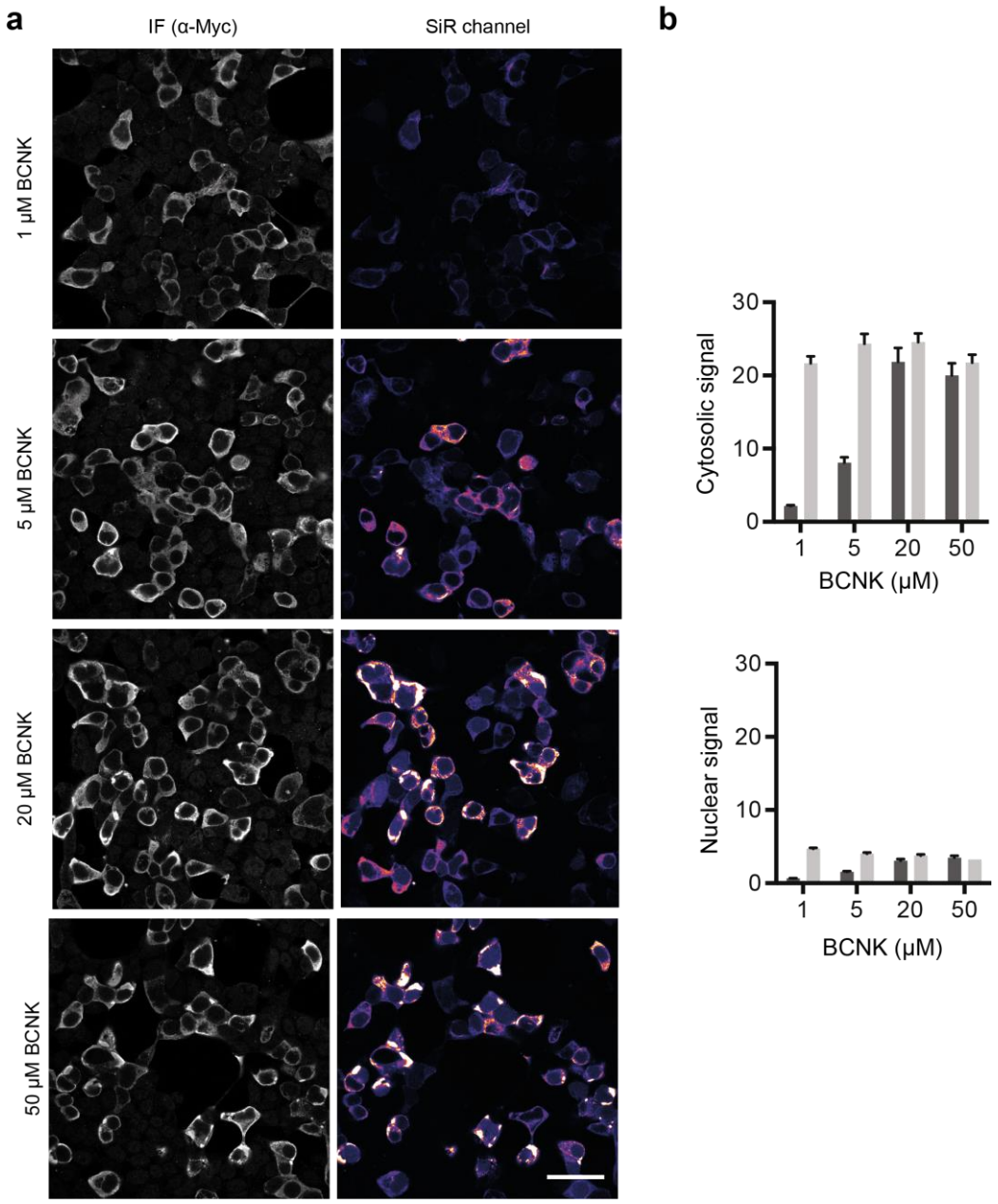

Figure 3.11. Reresentative full-size images shown in **Figure 3.9c** and used to generate data in **Figure 3.9d**. (a) HEK293 cells were transfected with constant amount of vimentin(N116TAG), PylRS-AF and $tRNA^{Pyl}_{CUA}$ genes and incubated in presence of varying amount of BCNK. The vimentin(116BCNK) was labellied in vivo by SiR-tetrazine (SiR channel) and post fixation and permeabilisation by anti-Myc antibody (IF (α-Myc)). Scalebar 50 µm. (b) Cells were manually segmented into nuclear and cytoplasmic regions and the signal in both BCNK and IF channel was quantified. Values shown are mean ± SEM. Each bar represents over 100 cells analysed.

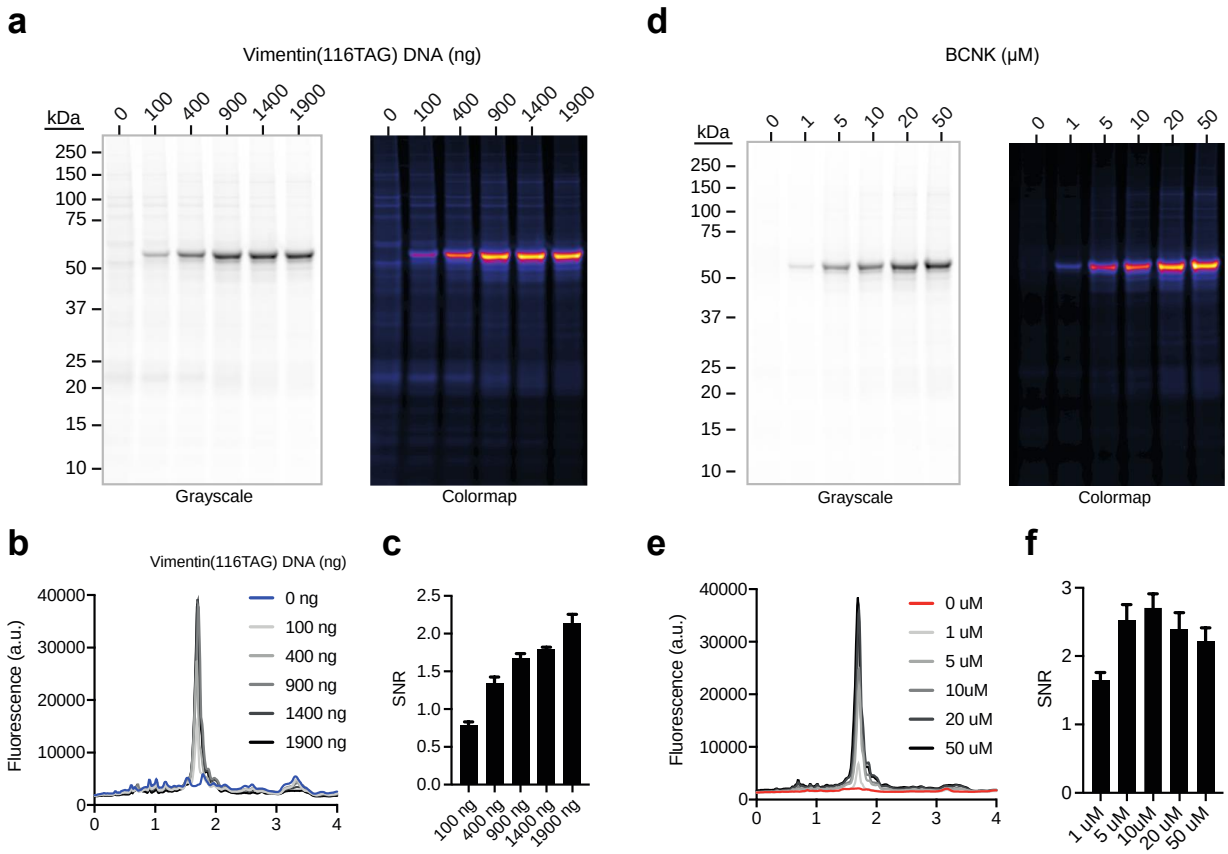

Figure 3.12. In-gel quantification of signal-to-noise ratio for in vivo labelling of vimentin via genetic code expansion. (a) Grayscale (left) and colourmap (right) fluorescent scans of SDS-PAGE gel of lysates of HEK293 cells transfected with increasing amount of Myc-Vimentin(116TAG) plasmid, incubated overnight in presence of 50 uM BCNK and labelled in vivo with SiR-tetrazine. (b) Fluorescent intensity traces corresponding to lanes in panel a. (c) Quantification of signal-to-noise ratio based on fluorescence intensity of the Myc-Vimentin(116BCNK) band and the rest of the proteome. (d) Grayscale (left) and colourmap (right) fluorescent scans of SDS-PAGE gel of lysates of HEK293 cells transfected with 1900 ng of Myc-Vimentin(116TAG) plasmid, incubated overnight in presence of varying concentrations of BCNK and labelled in vivo with SiR-tetrazine. (e) Fluorescent intensity traces corresponding to lanes in panel d. (f) Quantification of signal-to-noise ratio based on fluorescence intensity of the Myc-Vimentin(116BCNK) band and the rest of the proteome.

Next we wanted to confirm that decreasing the BCNK concentration provides control over expression levels of Myc-vimentin(N116BCNK) while maintaining SNR. As expected, we observed similar SNR for immunofluorescence labelling and BCNK labelling across different BCNK concentrations (**Figure 3.9c-d**). Interestingly, the cytosolic signal from SiR-tetrazine decreased correspondingly to the decreasing BCNK concentration, while immunostaining level remained relatively constant (**Figure 3.11b**). This result suggests that labelling proteins via unnatural amino acids allows quantitative analysis of protein levels via fluorescence imaging, overcoming the issue of low antibody penetration after fixation.

We conclude that the DNA concentration of the amber mutant transgene controls the SNR of the labelling method and that the observed SNR for the optimized conditions matches that of immunofluorescence, a widely used and accepted standard for fluorescent protein imaging (Stadler et al., 2013). In addition, the concentration of BCNK allows for direct control of the level of transgene expression without significantly compromising the SNR for fluorescent imaging.

3.4 The microprotein NoBody is rapidly exchanging component of P bodies

3.4.1 Small open reading frames

A group of proteins not amenable to standard labelling techniques altogether, are the recently discovered short open reading frame (sORF) encoded polypeptides (SEPs). sORFs are translated from traditionally designated untranslated regions or from splice variants of known protein coding sequences. Majority of sORFs contain less than 100 codons, coding for peptides with molecular weight below 10 kDa. To date, proteomic screens have reported hundreds of putative sORFs (Andrews and Rothnagel, 2014; Slavoff et al., 2013), a few of which have been confirmed to encode functional polypeptides (D'Lima et al., 2017). However, the functions of almost all SEPs in cells remain largely unknown, in part because it is difficult to tag individual SEPs for isolation, co-immunoprecipitation or live-cell imaging with traditional tagging tools,

while the use of peptide epitope tags limits tagging sites to protein termini, and cannot be extended to live-cell imaging.

3.4.2 Expression and site specific labelling of microprotein nobody

To demonstrate the advantage of the optimised protocol for genetic code expansion as a minimally perturbative method for live-cell imaging, we sought to describe the dynamic behaviour of a recently discovered microprotein NoBody (D'Lima et al., 2017). NoBody is a 7-kDa polypeptide that interacts with proteins involved in removal of 5'cap of mRNAs. Similarly to many proteins involved in the 5'-to-3' decay, NoBody localizes to cytosolic RNA-protein granules called P-bodies and its expression level is inversely correlated with number of P-bodies in cells (D'Lima et al., 2017), however, the information on its dynamics in living cells is missing.

When we transfected cells with increasing amount of Myc-NoBody(48TAG) gene, we observed strong increase in specific labelling and decrease of non-specific labelling of the proteome with increasing amount of transfected plasmid (**Figure 3.13**). In agreement with our previous experiments, expression of Myc-NoBody(48BCNK) showed a strong dependence on BCNK concentration, demonstrating the site specific incorporation of BCNK and allowing control over the expression level of NoBody by decreasing the BCNK concentration to minimal levels (**Figure 3.12**). We further confirmed the specificity of our labelling method *in situ* using immunofluorescence against Myc tag (**Figure 3.13a**).

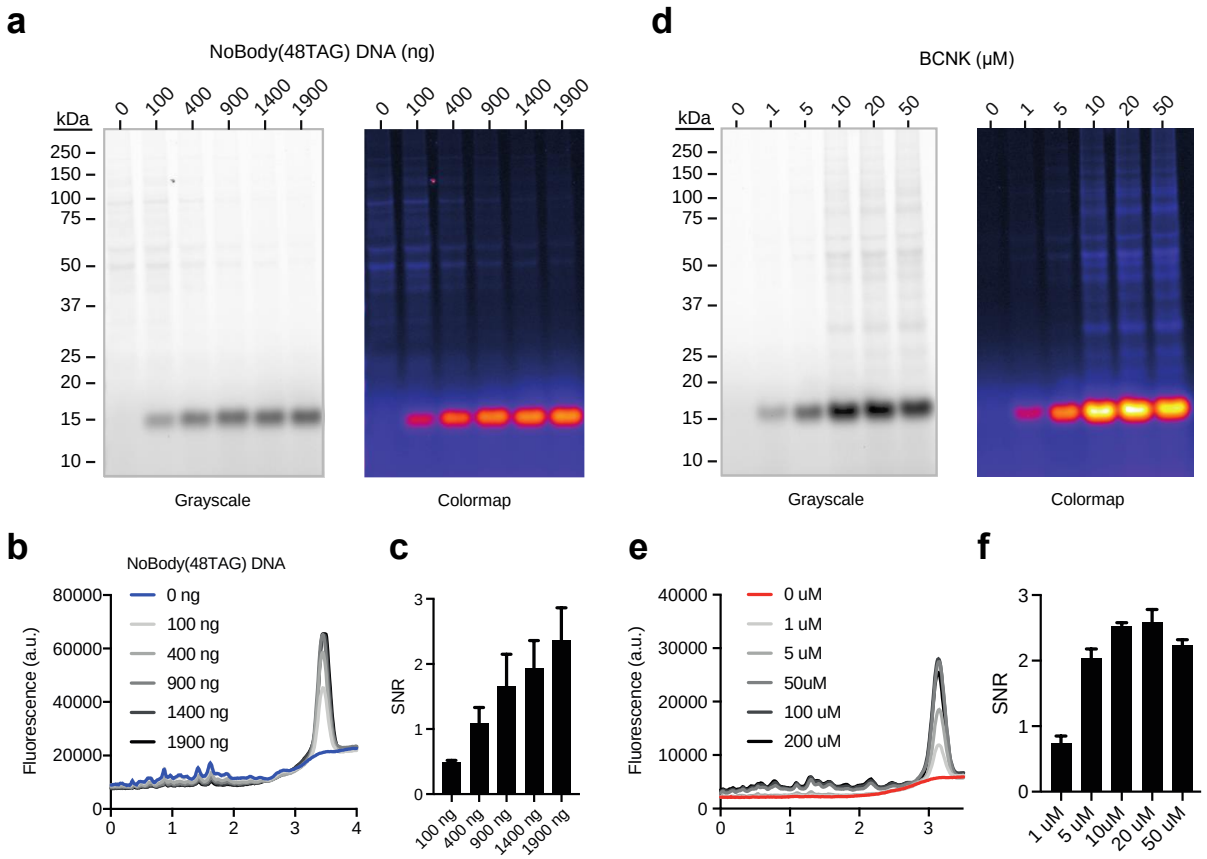

Figure 3.13. In-gel quantification of signal-to-noise ratio for in vivo labelling of NoBody via genetic code expansion. **(a)** Grayscale (left) and colourmap (right) fluorescent scans of SDS-PAGE gel of lysates of HEK293 cells transfected with increasing amount of Myc-NoBody(48TAG) plasmid, incubated overnight in presence of 50 uM BCNK and labelled in vivo with SiR-tetrazine. **(b)** Fluorescent intensity traces corresponding to lanes in panel a. **(c)** Quantification of signal-to-noise ratio based on fluorescence intensity of the Myc-NoBody(48BCNK) band and the rest of the proteome. **(d)** Grayscale (left) and colourmap (right) fluorescent scans of SDS-PAGE gel of lysates of HEK293 cells transfected with 400 ng of Myc-NoBody(48TAG) plasmid, incubated overnight in presence of varying concentrations of BCNK and labelled in vivo with SiR-tetrazine. **(e)** Fluorescent intensity traces corresponding to lanes in panel d. **(f)** Quantification of signal-to-noise ratio based on fluorescence intensity of the Myc-NoBody(48BCNK) band and the rest of the proteome.

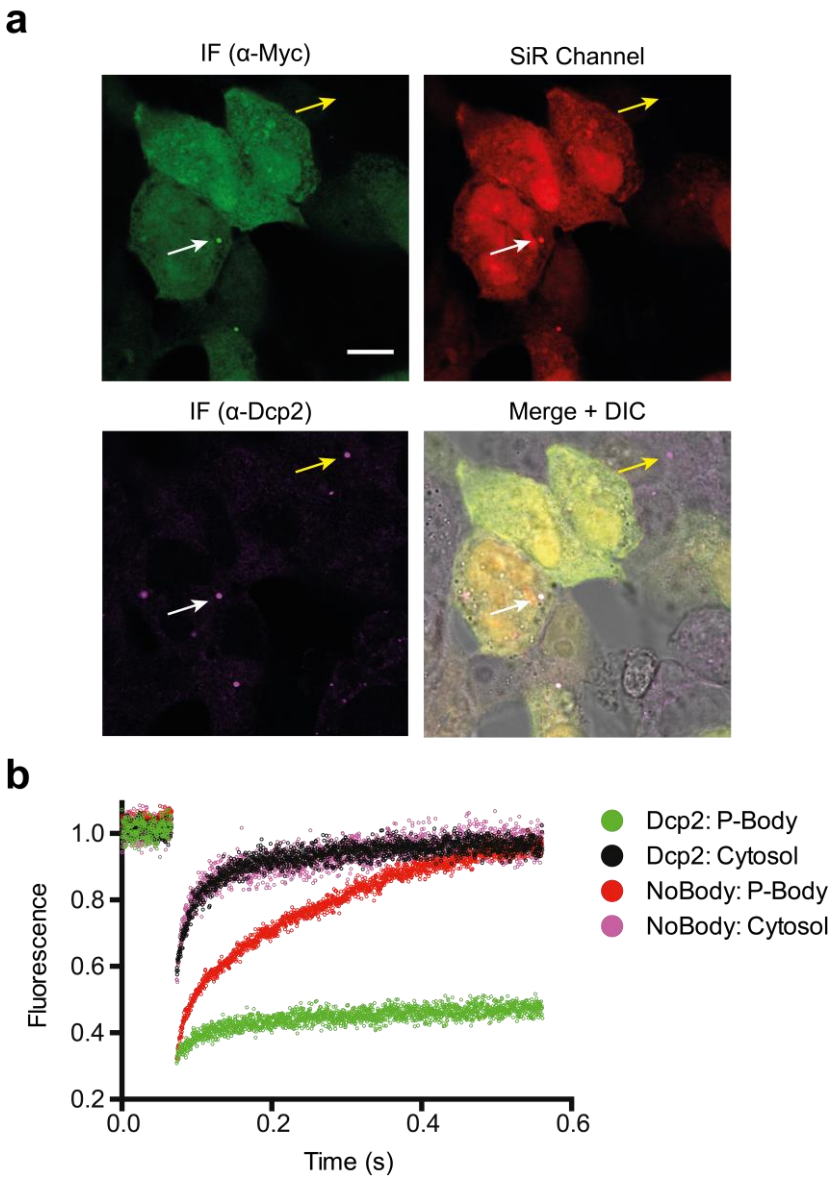

Figure 3.14. NoBody is a dynamic component of P-bodies. **(a)** Fluorescent images of fixed HEK293 cells transfected with Myc-NoBody(48TAG), MmPylRS-AF and PylT transgenes. Myc-NoBodyL48BCNK was labelled via α-Myc immunostaining (green) and SiR-tetrazine derivatization (red). Endogenous Dcp2 was labelled via immunostaining (magenta). **(b)** Representative curves of FRAP of Myc-NoBodyL48BCNK and Dcp2-GFP in P-body and in cytosol in live cells.

3.4.3 Live-cell dynamics of NoBody

Similarly to previous report, we found that only a small subset of cells expressing NoBody(48BCNK) contained P-bodies (**Figure 3.14a** and **Figure 3.15**), consistent with its role as a negative regulator of P-bodies. To limit overexpression, we used 5 µM BCNK – a concentration allowing us to observe some P-bodies in a subset of transfected cells – for all subsequent experiments.

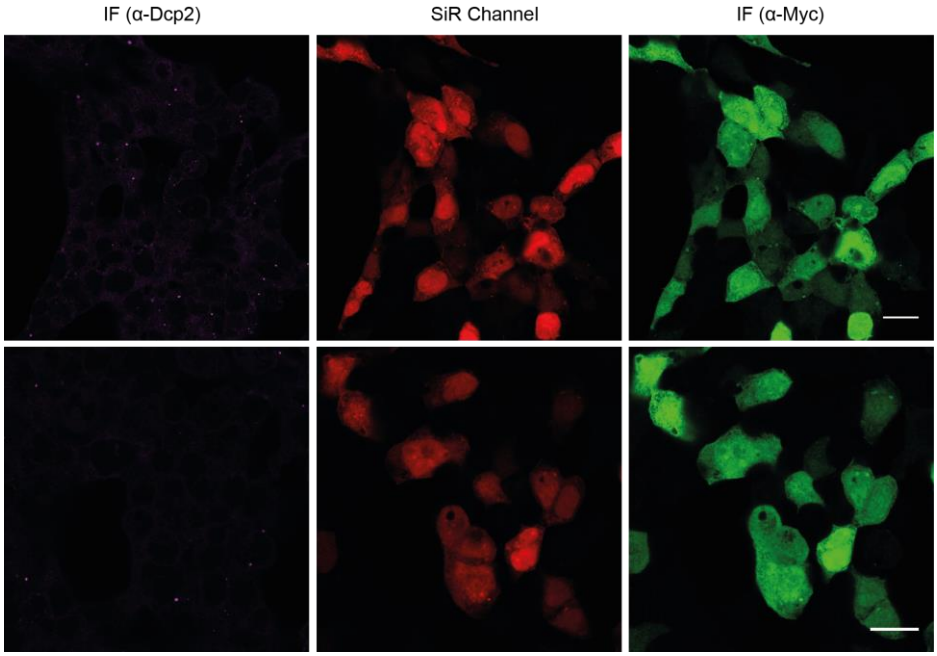

Figure 3.15. Multiple large-field-of-view images of fixed cells shown in **Figure 3.14**. Scalebar 20 µm

For live cell imaging of NoBody and its localisation to P-bodies, we used a GFP fusion of Dcp2, a core component of P-bodies (Franks et al., Mol Cell, 2009), as a P-body marker. We transfected HEK293 cells with *Mm*PylRS-AF, $tRNA^{Pyl}_{CUA}$ and NoBody(48TAG) transgenes, followed by transduction with lentivirus containing Dcp2-GFP fusion gene. To allow for easier observation of larger number of P-bodies, we stimulated cells with 0.3 mM sodium arsenite, inducing P-body assembly through oxidative stress (Kedersha et al., 2005). 1 hour after the sodium arsenite treatment, we were able to observe a large number of dynamic P-bodies cells, with all P-bodies enriched in the NoBody peptide (**Figure 3.16**)..

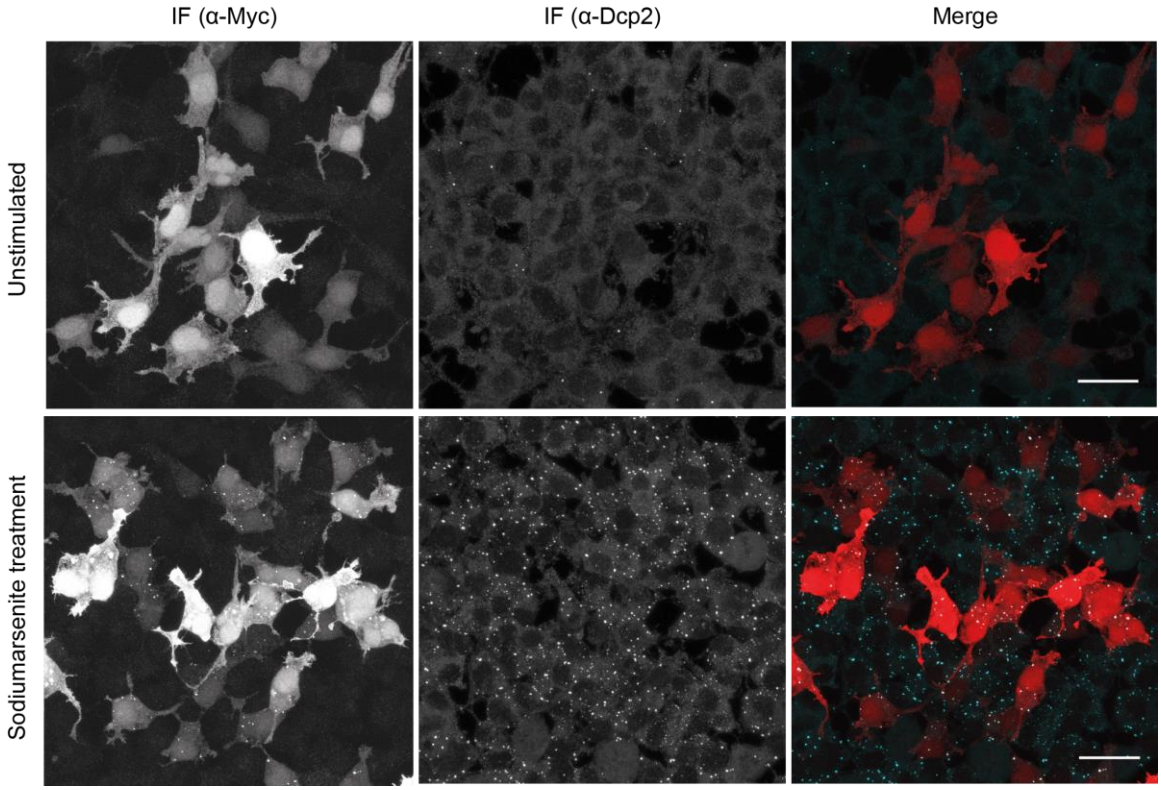

Figure 3.16. HEK 293 cells transduced with lentivirus containing Myc-NoBody transgene and stimulated with 0.3 mM sodium arsenite. Unstimulated cells show very low amount of P-bodies, especially in cells expressing Myc-NoBody. Scalebar 30 µm.

Finally, we wanted to examine whether NoBody is a stable component of P-bodies. Using Dcp2-GFP protein as a P-body marker, we measured fluorescence recovery after photobleaching (FRAP) to determine the rate of exchange of the two components to and from P-bodies. We observed that NoBody, unlike Dcp2, is rapidly exchanging with the cytosol, reaching complete recovery in less than 1 s, suggesting very transient interaction with components in the RNP granules (**Fig. 3.11b**).

3.5 Chapter discussion

The genetic code expansion in combination with bio-orthogonal chemistry provides a unique tool for protein labelling and imaging. The method allows for site-specific protein labelling in live cells with small, bright and photostable fluorophores but – despite the unique aspects of this approach – few key limits of the method exist. In this chapter we have identified and addressed these concerns.

We have demonstrated that by leveraging the recent developments in genetic code expansion in mammalian cells, we were able to achieve sufficient labelling to obtain super-resolved images of the mammalian cytoskeleton. This development should facilitate more precise imaging experiments due to the minimal size of the label and the linkage, taking full advantage of the spatial resolution of the imaging methods.

We further optimised the signal-to-noise ration of the labelling method by optimisation of the transfection protocol. The presence of endogenous amber codons poses a challenge to genetic code expansion via amber suppression. While the cell preferentially degrades most aberrantly extended proteins (Arribere et al., 2016) some do persist and are subsequently labelled by the fluorescent dye (**Figure 3.7**). To address this, we sought to shift the balance of amber codon readthrough towards the target protein by increasing the intracellular availability of the target mRNA. This approach led to over 5-fold increase in signal-to-noise ratio to a level comparable with immunofluorescence (**Figure 3.9**).

We used these improvements to image a small, 7 kDa polypeptide NoBody, leveraging the minimal perturbation caused by our labelling approach. Our results identify NoBody as a rapidly exchanging, dynamic component of P-bodies. The exact mechanism of P-body downregulation via P-body remains unclear (D'Lima et al., 2016). Our study suggests that direct structural perturbation is unlikely, given the transient nature of NoBody interaction. We note that this is a preliminary result and further experiments under varying conditions need to be carried out to confirm this hypothesis.

Combined together, the advances presented in this chapter provide a significant advancement in use of genetic code expansion for live-cell protein labelling. The minimal background seen

in our approach should allow for imaging of low abundance proteins in live cells. The imaging approach can be expanded to non-dividing primary cells, such as neurons and whole tissues when combined with the adeno-associated virus (AAVs) transduction system developed in our group (Ernst et al., 2016). This approach could be of particular interest for imaging neural micropeptides and microproteins (Makarewich and Olson, 2017), thanks to the minimal perturbation to target protein caused by genetic code expansion. Thanks to the mostly unrestrained position of the label within the polypeptide chain, this approach should allow for structural insights into macromolecular protein assemblies using particle averaging (Szymborska et al., 2013) and allow for development of novel FRET sensors (Baumdick et al., 2015).

3.6 Methods

3.6.1 Plasmid assembly

NoBody gene was cloned into previously reported plasmids for genetic code expansion in mammalian cells (Schmied et al., JACS 2014), using restriction cloning, introducing the amber codon at position 48 via quickchange to create pPB_NoBody48TAG_4xPylT. The second plasmid, pPB_MmPylRS-AF_4xPylT was cloned from previously reported plasmids using Gibson assembly (Uttamapinant et al., 2015).

3.6.2 Tissue culture

HEK293 cells (ATCC) were cultured in DMEM (ThermoFisher) media supplemented with 10% fetal bovine serum (FBS; ThermoFisher). Cells were passaged after reaching 80 – 90% confluence, by washing cells with PBS and trypsinization.

3.6.3 Flow cytometry

The HEK293 cells were transfected with appropriate plasmids using polyethyleneimine (MW ~40,000; Polysciences) in ratio 1 ug DNA to 4 ul of PEI (1 mg/ml solution in PBS). The cells were incubated for 1-2 days in presence or absence of the non-canonical amino acid. The cells were then trypsinized and analysed on FACS LSR-II (BD Biosciences) flow cytometer using the appropriate filter settings. The transfected cells (based on mCherry fluorescence threshold) were analysed, using mean GFP/mean mCherry ratio as the amber codon readthrough.

3.6.4 Lentivirus production

HEK293T cells (ATCC) were seeded into 10 cm dishes day prior. The cells were transfected at 90% confluence with combination of ENV plasmid, Packaging plasmid (pMDLg/p RRE) and pRSV-REV plasmid (Addgene #12251, #8454, #12253, respecively) and Dcp2-GFP plasmid using Lipfectamine 2000 mix (ThermoFisher). After 5 hours and 16 hours, the media

was exchanged for fresh media, starting the collection of the lentivirus. The media was collected 48 hours after transfection and again 72 hours post transfection, replacing with fresh media after first collection. The collected media was filtered using 0.45 µm filter to remove cellular debris. The media was centrifuged at 50,000 RCF for 150 minutes at 4C and the virus pellet was resuspended in ice cold PBS. For titration, HEK293 cells were transduced with serial dilution of the lentivirus and analyzed 2 days later by FACS.

3.6.5 Live cell labelling

HEK293 cells were seeded into poly-L-lysine coated 24-well plates of fibronectin (Merck) coated glass chambers. Upon reaching ~ 80% confluence, the media was exchanged for Optimem (Thermo Fischer) and transfection mixture of DNA and polyethyleneimine (MW ~40,000; Polysciences) in ratio 1 ug DNA to 4 ul of PEI (1 mg/ml solution in PBS) was added to the cells. After incubation for 4 hours at 37C, the media was exchanged for DMEM supplemented with 10% FBS and BCNK (Synnafix; solubilized in DMEM, adjusting pH by NaOH) of corresponding concentration. For live-cell imaging, cells were transduced with the appropriate lentivirus. Cells were grown for in presence of amino acid for 1-2 days. Before labelling, the cells were washed 2-3 times by DMEM media with 10% FBS to remove the excess BCNK. The cells were labelled for 20 minutes and labeled with 400 nM SiR-tetrazine (Spirochrome) and subsequently washed again with DMEM with 10% FBS. The imaging was carried out under 5% CO_2 atmosphere at 37C.

3.6.6 SDS-PAGE analysis

The cell media was removed by aspiration and cells were lysed with ice-cold RIPA buffer (25 mM Tris (pH 7.4), 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton X-100) supplemented with cOmplete protease inhibitors (Sigma-Aldrich). Cells were lysed on ice for 20 minutes and the lysates were cleared by centrifugation at 20,000 g at 4C. The supernatants were collected and protein concentration normalized using BCA assay (ThermoFisher). The lysates were analysed on 4-12% Bis-Tris gels (Novex) in LDS buffer supplemented with 5% β-mercaptoethanol. The gels were run at 200V for 35 – 45 minutes and subsequently analysed using Amersham Typhoon scanner (GE Life Sciences) with appropriate laser and filter settings.

3.6.7 Western blotting

For western blotting, the proteins were transferred onto PVDF membranes using the iBlot2 system, following the manufacturer's instructions. The membrane was blocked with 5% milk in TBS-T (20 mM Tris, 150 mM NaCl, 0.1 % Tween 20) for 1 hour at room temperature and then overnight at 4C with primary antibody α-GFP (Roche (13.1 and 7.1)), α-Myc (Cell Signaling (9B11)) or α-actin (Sigma-Aldrich (A3853)), using concentrations recommended by the manufacturer. The membrane was washed several times with TBS-T and incubated with corresponding HRP-conjugated secondary antibody (Cell Signalling) for 1 hour at room temperature. The membrane was washed few times with TBS-T and developed using SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher) according to manufacturer's instructions. Membrane images were acquired using a ChemiDoc system (Bio-Rad).

3.6.8 Live cell and fixed cell imaging

After labelling, the cells were imaged live or fixed using 4% paraformaldehyde in PBS for 10 mins at room temperature. The fixed cells were permeabilised using Triton X-100 and blocked using 3% bovine serum albumin (BSA) solution in PBS for 1 hour at room temperature. The cells were subsequently immunostained with primary antibodies α-Myc (Cell Signaling (9B11)) or α-Dcp2 (Abcam (ab28658)) for 1 hour at room temperature, washed 3 times with PBS and immunostained with fluorescently labeled secondary antibodies (Thermo Fisher). The cells were imaged using Leica SP8 or Leica STED confocal microscopes with appropriate laser and filter settings.

3.6.9 Fluorescence in situ hybridisation

HEK293T cells were transfected with the following plasmids: (U6- PylTU25C)4/EF1α-BCNKRS and (U6-PylTU25C)4/EF1α-vimentinN116TAG and incubated in the presence of 1 mM BCNK for 36-48 hours. The cells were subsequently washed and labelled with 400 nm tet1-CFDA. After washing of the excess dye, the cells were fixed using 4% formaldehyde in DPBS for 1 hr at room temperature and washed several times with DPBS. For in situ

hybridization, the cells were washed twice with 2X saline-sodium citrate (SSC) buffer with 50% v/v formamide (Ambion) and incubated for 30 minutes at 37°C in FISH buffer (10% dextran (Sigma), 2 mM vanadyl ribonucleoside complex (Sigma), 0.02% RNAse-free BSA (Life Technologies) and 1 µg/µl E. coli tRNA (Sigma) in 2X SSC, 50% formamide). Subsequently, 1.33 µg/µl of a DNA oligo probe (5'-GGAAACCCCGGGAATCTAACCCGGCTGAACGGATTTAGAG-3') with 5' terminal Alexa Fluor 647 was mixed to the buffer and hybridized overnight at 37°C. Before imaging, the cells were washed twice in 2X SSC, 50% v/v formamide for 30 minutes at 37°C and once with DPBS.

Chapter 4 An Evolved *M. alvus* pyrrolysyl-tRNA synthetase/tRNA pair is highly active and orthogonal in mammalian cells.

4.1 Chapter Introduction

The efforts to incorporate non-canonical amino acids into proteins in mammalian cells rely on the availability of new orthogonal aaRS/tRNA pairs. To date, only a limited number of orthogonal aaRS/tRNA pairs in mammalian cells have been reported (Italia et al., 2017). With the exception of the *Methanococcus mazei* (*Mm*) and *barkeri* pyrrolysine-tRNA synthetase/tRNA pairs, the majority are orthogonal in eukaryotes but not in *E. coli*. This chapter demonstrates that the recently engineered PylRS/tRNA$^{Pyl}(6)_{CUA}$ pair from *Methanomethylophilus alvus* (*Ma*) is highly active and orthogonal in mammalian cells, making it only second pair orthogonal in both eukaryotes and wild-type *E. coli*. In addition, we demonstrate the *Ma*PylRS/tRNA$^{Pyl}(6)_{CUA}$ pair is orthogonal to the widely used *Mm*PylRS/tRNA$^{Pyl}_{CUA}$ pair and that these two pairs can be used together to site-specifically incorporate two distinct non-canonical amino acids into a single protein. This development paves a way for an efficient incorporation of multiple distinct amino acids into proteins in mammalian cells.

The following section has been accepted for publication as:

Beránek, V., Willis, J.C.W, Chin, J.W., 2018 An evolved *M. alvus* pyrrolysyl-tRNA synthetase/tRNA pair is highly active and orthogonal in mammalian cells. Biochemistry, *In press*.

The following list specifies my contributions to the manuscript as well as the contributions of the other co-authors:

J.W.C. defined the direction of research. V.B. and J.W.C. designed the experiments. V.B. performed the experiments and analysed the data. J.C.W.W. identified the *Ma*PylRS mutant used for double incorporation. V.B. and J.W.C. wrote the paper.

Supplementary figures were inserted into the text where first mentioned and the numbering of figures has been modified from the manuscript.

4.2 An evolved *M. alvus* pyrrolysyl-tRNA synthetase/tRNA pair is highly active and orthogonal in mammalian cells.

Genetically encoding the site-specific co-translational incorporation of non-canonical amino acids into proteins in eukaryotic cells and animals has provided numerous strategies for imaging and controlling the functions of proteins in their native environment (Chin, 2014). Extensions of these approaches have enabled the tagging and labelling of cell-specific proteomes via stochastic orthogonal recoding of translation (SORT) (Elliott et al., 2016; Elliott et al., 2014; Hoffmann et al., 2018; Krogager et al., 2018).

The incorporation of ncAAs into proteins relies on the development of orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA pairs: the orthogonal aaRS selectively recognizes its cognate orthogonal tRNA over endogenous tRNAs and the orthogonal tRNA is a substrate for the orthogonal aaRS, but a poor substrate for endogenous synthetases. Since the sets of endogenous synthetases and tRNAs differ between organisms, aaRS pairs that are orthogonal in one system are commonly not orthogonal in another (Chin, 2014). For example, the *Methanocaldococcus janaschii* (Mj) TyrRS pair that has been extensively used for genetic code expansion in *E. coli* (*Ec*) cannot be used for genetic code expansion in eukaryotic cells because it is not orthogonal with respect to endogenous eukaryotic aaRS/tRNA pairs.

The *Ec*TyrRS/^{Tyr}tRNA (Chin et al., 2003; Sakamoto et al., 2002) pair and *Ec*LeuRS/^{Leu}tRNA pair(Wu et al., 2004) are orthogonal in eukaryotic cells and variants of these pairs have been discovered, primarily by directed evolution in yeast or subsequent screening, that enable the incorporation of a range of ncAAs in eukaryotic systems. The PylRS/^{Pyl}tRNA pair from *M*m is commonly considered an ideal pair for genetic code expansion because it is orthogonal in both *E. coli* and eukaryotic cells and animals (Chin, 2014). This has facilitated the discovery and characterization of *Mm*PylRS variants that incorporate ncAAs in *E. coli* and the transfer of these variant to eukaryotic systems, thereby facilitating genetic code expansion in eukaryotic cells and animals. Recent work has demonstrated that SepRS/^{Sep}tRNA$_{CUA}$ pair that has been evolved for efficient incorporation of phosphoserine and its nonhydrolyzable analog (Rogerson et al., 2015), can be further evolved to incorporate phosphothreonine in *E. coli* (Zhang et al., 2017) and is also orthogonal in mammalian cells (Beranek et al., 2018). The ability to

incorporate ncAAs into proteins in mammalian cells has been further expanded by strategies that replace the genomically encoded *Ec*TrpRS/^{Trp}tRNA pair in *E. coli* with the *Sacchromyces cerevisiae* (*Sc)* TrpRS/^{Trp}tRNA pair (Hughes and Ellington, 2010; Iraha et al., 2010; Italia et al., 2017). Since the *Sc*TrpRS/^{Trp}tRNA is orthogonal in *E. coli*, suppressor derivatives of the *Ec*TrpRS/^{Trp}tRNA pair can be introduced into the resulting *E. coli* strains and evolved for ncAA incorporation. The resulting *Ec*TrpRS/^{Trp}tRNA pairs can then be used for genetic code expansion in mammalian cells, where they are orthogonal. Recent work has extended this strategy to the *Ec*TyrRS/ ^{Tyr}tRNA pair (Italia et al., 2018). In several cases, aaRS/tRNA pairs that are active and orthogonal with respect to the endogenous aaRSs and tRNAs in mammalian cells have also been shown to be orthogonal with respect to other orthogonal pairs; creating 'mutually orthogonal' pairs in mammalian cells (Xiao et al., 2013; Zheng et al., 2017a; Zheng et al., 2017b).

Willis and Chin recently discovered that a new class of PylRS/^{Pyl}tRNA pairs from *Methanomassiliicocales* are active and orthogonal in *E. coli*.(Borrel et al., 2014; Willis and Chin, 2018) These pairs, unlike the *Mm*PylRS/Mm^{Pyl}tRNA pair, lack the N-terminal domain of PylRS, which was previously thought to be essential for tRNA recognition and aminoacylation (Herring et al., 2007; Jiang and Krzycki, 2012). We showed that certain orthogonal pairs from this class are naturally mutually orthogonal to the *Mm*PylRS (Willis and Chin, 2018); this surprising result demonstrated that there is sufficient divergence between archeal PylRS/^{Pyl}tRNA pairs to generate mutually orthogonal pairs for the same amino acid within a domain of life. We developed a number of exceptionally active and orthogonal pairs, including the *Ma*PylRS/*Ma*^{Pyl}tRNA(6) pair, that are mutually orthogonal to the *Mm*PylRS/*Mm*^{Pyl}tRNA pair by virtue of mutations introduced into the body of *Ma*^{Pyl}tRNA (Willis and Chin, 2018). Moreover, we showed that, since the active sites of *Mm*PylRS and *Ma*PylRS share common substrate recognition determinants, we could transplant mutations that direct the selective incorporation of specific ncAAs from the *Mm*PylRS active site to the *Ma*PylRS active site to reprogram its substrate specificity (Willis and Chin, 2018). Finally, we showed that by diverging the active sites of *Ma*PylRS and *Mm*PylRS to selectively recognize distinct substrates and altering the anticodons of *Ma*^{Pyl}tRNA(6) and *Mm*^{Pyl}tRNA to decode

distinct codons in orthogonal translation we could use these pairs in the same cell to direct the incorporation of two distinct ncAAs into a single polypeptide (Willis and Chin, 2018). Here we show that the *Ma*PylRS/*Ma*^{Pyl}tRNA(6) is highly active and orthogonal in mammalian cells, where it is also mutually orthogonal to the *Mm*PylRS/*Mm*^{Pyl}tRNA (**Figure 4.1**), and that derivatives of the two pairs can be used to incorporate distinct amino acids into a protein in mammalian cells.

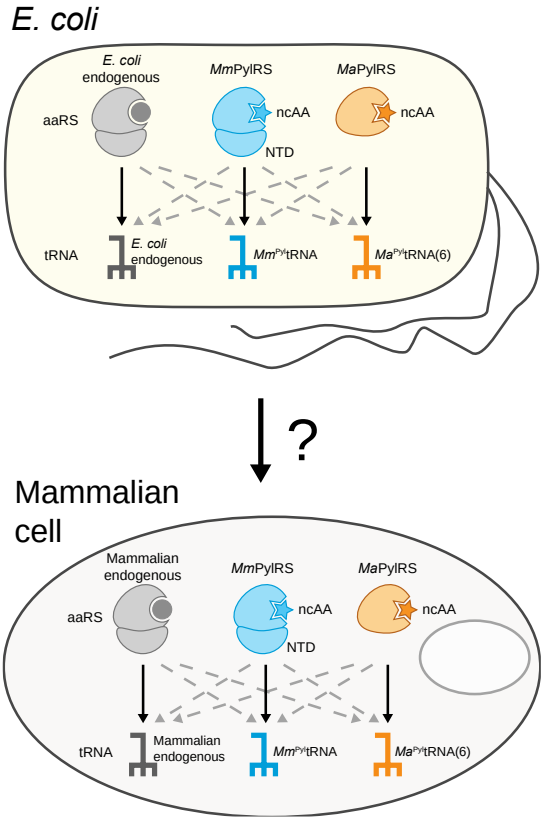

Figure 4.1. The evolved *Ma*PylRS/*Ma*^{Pyl}tRNA(6) is orthogonal with respect to the endogenous *E. coli* aaRS/tRNA pairs as well as mutually orthogonal to the *Mm*PylRS/*Mm*^{Pyl}tRNA(6) pair. Here we ask if this pair is also orthogonal and mutually orthogonal in mammalian cells. NTD is N-terminal domain.

We first demonstrated that the *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pair is active in mammalian cells, and that both *Ma*PylRS and *Ma*^{Pyl}tRNA(6)$_{CUA}$ are orthogonal with respect to the endogenous

tRNAs and the aaRSs in human cells. To this aim, we cloned coding sequences of *Ma*PylRS and *Ma*^{Pyl}tRNA(6)$_{CUA}$ into a vector for mammalian expression;(Schmied et al., 2014) we cloned *Ma*PylRS under the EF1a promoter and four copies of the *Ma*^{Pyl}tRNA(6)$_{CUA}$ under the human U6 promoter into the same vector. We transiently co-transfected HEK293 cells with the *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ vector and an mCherry-TAG-GFP reporter (Beranek et al., 2018) and cultured the resulting cells in presence or absence of Nε–((*tert*-butoxy)carbonyl)-L-lysine (BocK), a known substrate for this pair.(Willis and Chin, 2018) We measured the ratio of GFP/mCherry fluorescence by flow cytometry (**Figure 4.2a**, **Figure 4.3**) and fluorescence microscopy (**Figure 4.4**), and performed control experiments with the well-characterised *Mm*PylRS/^{Pyl}tRNA$_{CUA}$ pair.

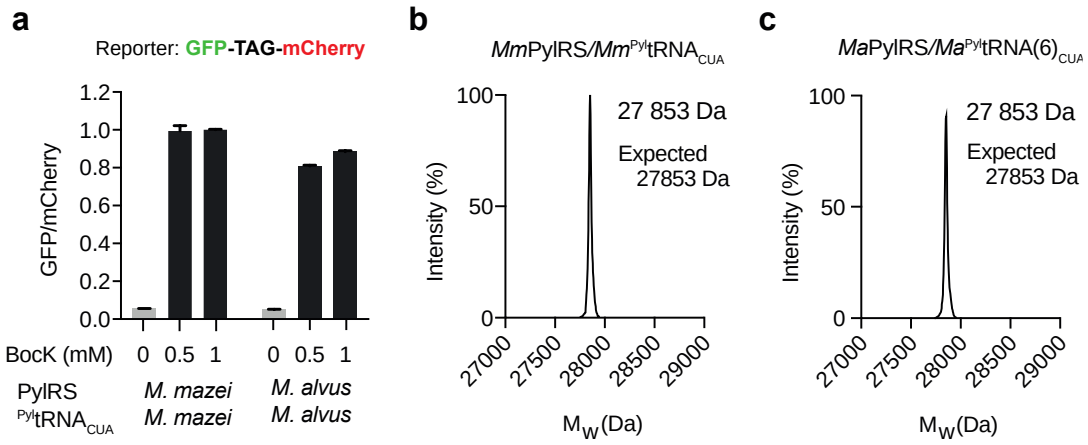

Figure 4.2. *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ is active and orthogonal in mammalian cells. (a) *Ma*PylRS and *Mm*PylRS show comparable, BocK dependent, readthrough of the amber stop codon with their cognate tRNAs. Data represent mean +/- SD from two biological replicates. (b) ESI-MS of purified sfGFP confirms quantitative incorporation of BocK via the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair. (c) ESI-MS of purified sfGFP confirms quantitative incorporation of BocK via the *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pair.

We observed minimal readthrough of the amber codon by the *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pair in the absence of BocK (**Figure 4.2a)**. This demonstrates that *Ma*^{Pyl}tRNA(6)$_{CUA}$ is orthogonal with respect to the aminoacyl-tRNA synthetases that are endogenous in human cells. Upon

addition of 0.5 mM and 1 mM BocK, we observed substantial readthrough of the TAG codon by the *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pair. The level of amber codon readthrough mediated by the *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pair is comparable to that mediated by the highly active *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair (**Figure 4.2a**); this demonstrates that the *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pair is highly active in mammalian cells.

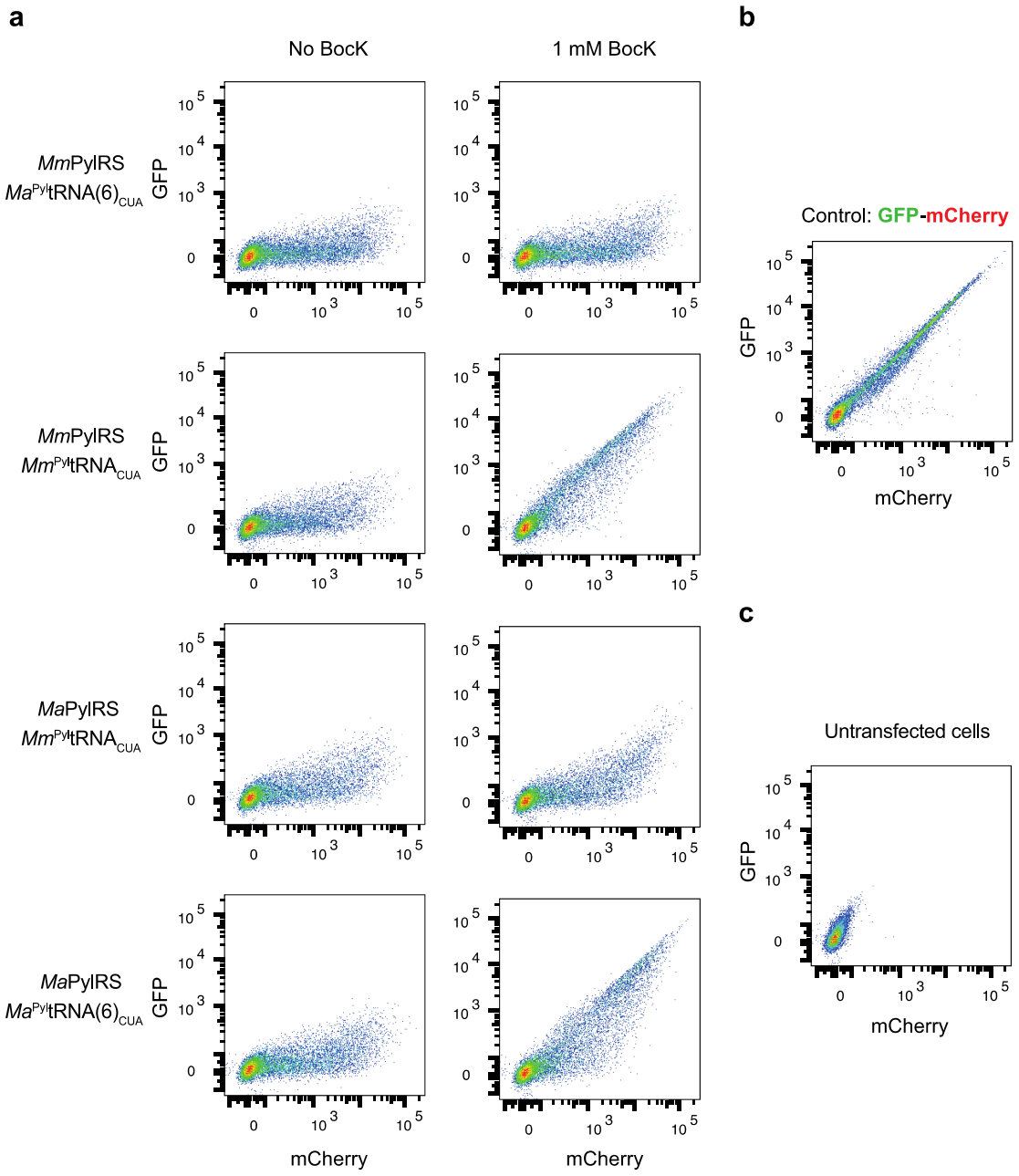

Figure 4.3. *Ma*PylRS/Ma^{Pyl}tRNA(6)$_{CUA}$ is active and orthogonal in mammalian cells and mutually orthogonal to *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$. (a) Flow cytometry data summarised in **Figures 4.2** and **4.5**. *Ma*PylRS and *Mm*PylRS show comparable, BocK dependent readthrough of the amber stop codon with their cognate tRNAs but not the non-cognate tRNAs. Readthrough control (b) and untransfected cells (c) are shown.

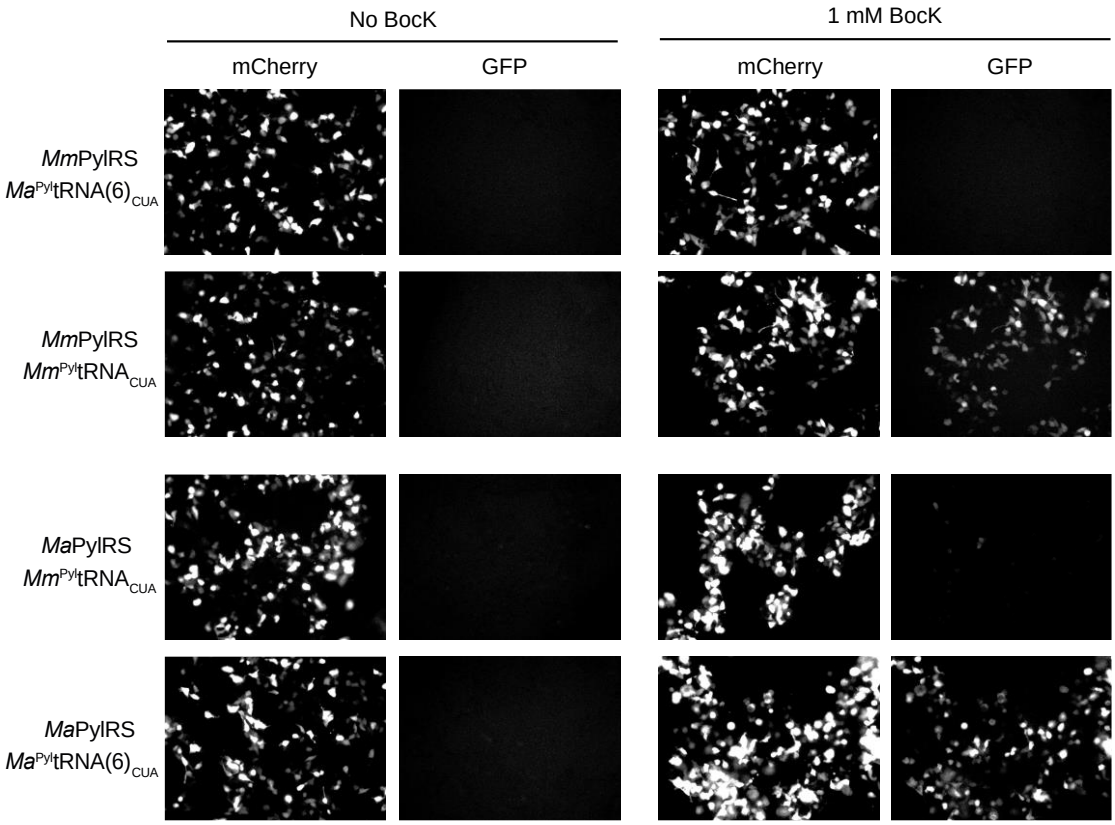

Figure 4.4. *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ is active and orthogonal in mammalian cells and mutually orthogonal to *Mm*PylRS/*Mm*tRNA$^{Pyl}_{CUA}$. Fluorescence microscopy data summarised in **Figures 4.2** and **4.5**. *Ma*PylRS and *Mm*PylRS show comparable, BocK dependent readthrough of the amber stop codon with their cognate tRNAs but not the non-cognate tRNAs.

To demonstrate that *Ma*PylRS is functionally orthogonal in human cells we co-transfected plasmids encoding the *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pair and *GFP(150TAG)His$_6$* and cultured the cells in the presence of 1 mM BocK. ESI-MS of the resulting GFP has the expected mass (**Figure 4.2b**) and is indistinguishable from a control in which we used the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair to incorporate BocK in to *GFP(150TAG)His$_6$* (**Figure 4.2c**). Taken together our experiments reveal that the *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pair is a highly active and orthogonal pair in mammalian cells.

Next, we aimed to demonstrate that the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair and *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pair are mutually orthogonal in their aminoacylation specificity when expressed in mammalian cells. We swapped the 4x U6-tRNA cassette between the *Mm* and *Ma* expression vectors producing a plasmid containing *Mm*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pair and a plasmid containing *Ma*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair. We co-transfected the mCherry-TAG-GFP reporter with each of the two and cultured the cells in presence and absence of BocK. We compared the readthrough of the amber stop codon by these non-cognate pairs to that mediated by the *Mm* and *Ma* derived cognate pairs (**Figure 4.5**). Our data show that the non-cognate *Mm*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ and *Ma*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pairs lead to minimal read through of the amber stop codon (**Figure 4.5**), while the cognate pairs lead to efficient amber suppression (**Figure 4.2**). This demonstrates that these two PylRS/^{Pyl}tRNA$_{CUA}$ pairs are mutually orthogonal in mammalian cells.

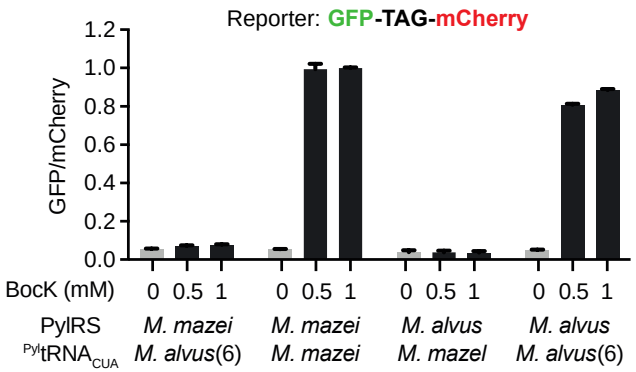

Figure 4.5. *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ is mutually orthogonal to *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair in mammalian cells. *Ma*PylRS and *Mm*PylRS show comparable, BocK dependent aminoacylation of their cognate tRNA and minimal cross-aminoacylation of the non-cognate ^{Pyl}tRNA$_{CUA}$. Data represent mean +/- SD from two biological replicates.

Next we differentiated the active sites of *Ma*PylRS and *Mm*PylRS such that they selectively recognize distinct substrates. By screening a collection of *Ma*PylRS mutants for non-natural substrate specificity in *E. coli*, we discovered a variant of *Ma*PylRS, *Ma*PylRS(mut) that incorporates 3-methyl-L-histidine (Me-His) but not BocK, this synthetase contains the

mutations L121M, L125I, Y126F, M129A and V168F. And we find that *Mm*PylRS directs the incorporation of BocK but not Me-His, these specificities are maintained in mammalian cells (**Figure 4.6a**). Finally we performed a double incorporation of BocK and Me-His into GFP(101TGA,150TAG) using the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CAU}$ and *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pairs. Production of full-length protein was dependent on addition of both ncAAs (**Figure 4.6b**), consistent with the site-specific incorporation of both amino acids into GFP in mammalian cells.

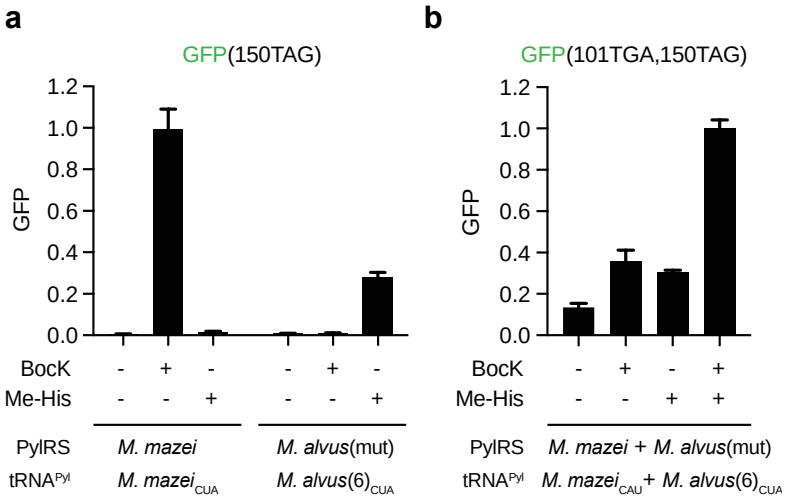

Figure 4.6. (a) *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ and *Ma*PylRS(mut)/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pairs selectively incorporate BocK and Me-His in response to amber stop codon. Data represent mean +/- SD from three biological replicates. All ncAAs were used at 1 mM. (b) Double incorporation of Me-His and BocK into GFP(101TGA,150TAG) using the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CAU}$ and *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pairs. Data represent mean +/- SD from three biological replicates. All amino acids were used at 1 mM.

We have demonstrated that the *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pair is orthogonal with respect to the synthetases and tRNAs present in mammalian cells, and is highly active in mammalian cells. As the *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pair is also orthogonal in *E. coli*, and we have demonstrated that mutations discovered in the active site of *Mm*PylRS or *Mb*PylRS can be transplanted into *Ma*PylRS to reprogram its substrate specificity (Willis and Chin, 2018), it

will be possible to rapidly expand the substrate scope of the *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pair. Moreover, as the *Ma*PylRS is a single domain protein that lacks the poorly soluble N-terminal domain of *Mm*PylRS it may be even more amenable to directed evolution in *E. coli* than the *Mm*PylRS system. Our results demonstrate that *Ma*PylRS mutants discovered and characterized in *E. coli* will be of direct utility in mammalian cells.

Finally we have confirmed that the mutual orthogonality of the *Mm*PylRS/*Mm*^{Pyl}tRNA pair and *Ma*PylRS/*Ma*^{Pyl}tRNA(6) pair, that we have characterized in *E. coli*, is maintained in mammalian cells and shown that these pairs can be used together for (un-optimized) double incorporation. These are therefore the first mutually orthogonal pairs in which each pair is itself orthogonal in both *E. coli* and eukaryotic systems. We anticipate that this foundational advance will facilitate multiplexed proteome labelling (Elliott et al., 2016; Elliott et al., 2014; Hoffmann et al., 2018; Krogager et al., 2018). Combinations of the advances reported herein, together with strategies for creating additional blank codons and increasing the efficiency of multi-site ncAA incorporation in mammalian cells may facilitate the site-specific incorporation of diverse, and currently inaccessible, combinations of ncAAs into proteins in mammalian cells (Xiao et al., 2013; Zheng et al., 2017a; Zheng et al., 2017b).

4.3 References

Beranek, V., Reinkemeier, C.D., Zhang, M.S., Liang, A.D., Kym, G., and Chin, J.W. (2018). Genetically Encoded Protein Phosphorylation in Mammalian Cells. Cell Chem Biol 25, 1–8.

Borrel, G., Gaci, N., Peyret, P., O'Toole, P.W., Gribaldo, S., and Brugere, J.F. (2014). Unique characteristics of the pyrrolysine system in the 7th order of methanogens: implications for the evolution of a genetic code expansion cassette. Archaea 2014, 374146.

Chin, J.W. (2014). Expanding and reprogramming the genetic code of cells and animals. Annu Rev Biochem 83, 379-408.

Chin, J.W., Cropp, T.A., Anderson, J.C., Mukherji, M., Zhang, Z., and Schultz, P.G. (2003). An expanded eukaryotic genetic code. Science 301, 964-967.

Elliott, T.S., Bianco, A., Townsley, F.M., Fried, S.D., and Chin, J.W. (2016). Tagging and Enriching Proteins Enables Cell-Specific Proteomics. Cell Chem Biol 23, 805-815.

Elliott, T.S., Townsley, F.M., Bianco, A., Ernst, R.J., Sachdeva, A., Elsässer, S.J., Davis, L., Lang, K., Pisa, R., and Greiss, S. (2014). Proteome labelling and protein identification in specific tissues and at specific developmental stages in an animal. Nat Biotechnol 32, 465.

Herring, S., Ambrogelly, A., Gundllapalli, S., O'Donoghue, P., Polycarpo, C.R., and Söll, D. (2007). The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity. FEBS Lett 581, 3197-3203.

Hoffmann, J.E., Dziuba, D., Stein, F., and Schultz, C. (2018). A Bifunctional Noncanonical Amino Acid: Synthesis, Expression, and Residue-Specific Proteome-wide Incorporation. Biochemistry 57, 4747-4752.

Hughes, R.A., and Ellington, A.D. (2010). Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA. Nucleic Acids Res 38, 6813-6830.

Iraha, F., Oki, K., Kobayashi, T., Ohno, S., Yokogawa, T., Nishikawa, K., Yokoyama, S., and Sakamoto, K. (2010). Functional replacement of the endogenous tyrosyl-tRNA synthetase-tRNATyr pair by the archaeal tyrosine pair in Escherichia coli for genetic code expansion. Nucleic Acids Res 38, 3682-3691.

Italia, J.S., Addy, P.S., Wrobel, C.J., Crawford, L.A., Lajoie, M.J., Zheng, Y., and Chatterjee, A. (2017). An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. Nat Chem Biol 13, 446.

Italia, J.S., Zheng, Y., Kelemen, R.E., Erickson, S.B., Addy, P.S. and Chatterjee, A., 2017. Expanding the genetic code of mammalian cells. Biochemical Society Transactions, 45(2), pp.555-562.

Italia, J.S., Latour, C., Wrobel, C.J., and Chatterjee, A. (2018). Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes. Cell Chem Biol.

Jiang, R., and Krzycki, J.A. (2012). PylSn and the homologous N-terminal domain of pyrrolysyl-tRNA synthetase bind the tRNA that is essential for the genetic encoding of pyrrolysine. J Biol Chem, jbc. M112. 396754.

Krogager, T.P., Ernst, R.J., Elliott, T.S., Calo, L., Beránek, V., Ciabatti, E., Spillantini, M.G., Tripodi, M., Hastings, M.H., and Chin, J.W. (2018). Labelling and identifying cell-specific proteomes in the mouse brain. Nat Biotechnol 36, 156.

Rogerson, D.T., Sachdeva, A., Wang, K., Haq, T., Kazlauskaite, A., Hancock, S.M., Huguenin-Dezot, N., Muqit, M.M., Fry, A.M., Bayliss, R., et al. (2015). Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. Nat Chem Biol 11, 496-503.

Sakamoto, K., Hayashi, A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., Kobayashi, T., Kitabatake, M., Takio, K., and Saito, K. (2002). Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. Nucleic Acids Res 30, 4692-4699.

Schmied, W.H., Elsasser, S.J., Uttamapinant, C., and Chin, J.W. (2014). Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1. J Am Chem Soc 136, 15577-15583.

Willis, J.C.W., and Chin, J.W. (2018). Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. Nat Chem 10, 831-837.

Wu, N., Deiters, A., Cropp, T.A., King, D., and Schultz, P.G. (2004). A genetically encoded photocaged amino acid. J Am Chem Soc 126, 14306-14307.

Xiao, H., Chatterjee, A., Choi, S.h., Bajjuri, K.M., Sinha, S.C., and Schultz, P.G. (2013). Genetic incorporation of multiple unnatural amino acids into proteins in mammalian cells. Angew Chem, Int Ed 52, 14080-14083.

Zhang, M.S., Brunner, S.F., Huguenin-Dezot, N., Liang, A.D., Schmied, W.H., Rogerson, D.T., and Chin, J.W. (2017). Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. Nat Methods 14, 729.

Zheng, Y., Addy, P.S., Mukherjee, R., and Chatterjee, A. (2017a). Defining the current scope and limitations of dual noncanonical amino acid mutagenesis in mammalian cells. Chem Sci 8, 7211-7217.

Zheng, Y., Mukherjee, R., Chin, M.A., Igo, P., Gilgenast, M.J., and Chatterjee, A. (2017b). Expanding the Scope of Single-and Double-Noncanonical Amino Acid Mutagenesis in Mammalian Cells Using Orthogonal Polyspecific Leucyl-tRNA Synthetases. Biochemistry 57, 441-445.

4.4 Methods

4.4.1 Plasmid assembly

*Ma*PylRS coding sequence was synthesised as double stranded DNA fragment (Integrated DNA Technologies) and introduced into the previously reported plasmid for mammalian expression(Schmied et al., 2014) via restriction cloning, creating the pPB_MaPylRS_4xU6-Mm^{Pyl}tRNA$_{CUA}$ plasmid. The *Ma*^{Pyl}tRNA(6)$_{CUA}$ gene under the U6 promoter and followed by terminator was synthesized as double stranded DNA fragment (Integrated DNA Technologies) with the following sequence (*Ma*^{Pyl}tRNA(6)$_{CUA}$ gene underlined):

CCTAGTTGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATA
CGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAG
ATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAG
TTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGGGGACGGT
CCGGCGACCAGCGGGTCTCTAAAACCTAGCATAGCGGGGTTCGACACCCCGGTC
TCTCGGACAAGTGCGGTTTTT

This fragment was PCR amplified in four reactions with 4 pairs of mutually overlapping primers and assembled via Gibson assembly with the backbone of pPB_MaPylRS_4xU6-Mm^{Pyl}tRNA$_{CUA}$ plasmid digested with StuI and AsiSI restriction enzymes, replacing the 4xU6-Mm^{Pyl}tRNA$_{CUA}$ cassette and creating the pPB_MaPylRS_4xU6-Ma^{Pyl}tRNA$_{CUA}$ plasmid. The plasmid pPB_MaPylRS_4xU6-Ma^{Pyl}tRNA$_{CAU}$ and pPB_MmPylRS_4xU6-Mm^{Pyl}tRNA$_{CAU}$ were constructed analogously.

The pPB_MmPylRS_4xU6-Ma^{Pyl}tRNA$_{CUA}$ and pPB_GFP150TAG_4xU6-Ma^{Pyl}tRNA$_{CUA}$ plasmids were created by restriction cloning from the previously reported pPB_MmPylRS_4xU6-Mm^{Pyl}tRNA$_{CUA}$ and pPB_GFP150TAG_4xU6-Mm^{Pyl}tRNA$_{CUA}$(Schmied et al., 2014), using the StuI and AsiSI restriction sites. Lastly, the pPB_ GFP(101TGA,150TAG) and pPB_MaPylRS(mut)_4xU6-Ma^{Pyl}tRNA$_{CUA}$, were created using Gibson cloning from the

above described plasmids. The GFP150TAG encodes all of the 20 canonical amino acids, using the following codons:

Phe (TTC(12)), Leu (CTC(2), CTG(18)), Ile (ATC (11)), Met (ATG (5)), Val (GTG (19)), Ser (TCT (1), TCC (1), AGC (9)), Pro (CCT (2), CCC (8)), Thr (ACC (16), ACA (2)), Ala (GCT (1), GCC (7)), Tyr (TAG (1)), His (CAT (4), CAC (12)), Gln (CAG (7)), Asn (AAT (1), ACC (11)), Lys (AAA (1), AAG (19)), Asp (GAT (4), GAC (14)), Glu (GAA (7), GAG (9)), Cys (TGC (2)), Trp (TGG (1)), Arg (CGG (6), AGA (2)), Gly (GGC (21), GAA (2))

4.4.2 Tissue culture

HEK293 (female human embryonic kidney) cells (ATCC) were cultured in Dulbecco's Modifid Eagle Medium (DMEM) (Gibco) supplemented with 10% fetal bovine serum (FBS) (Gibco) at 37°C and 5% CO_2 in humidified incubator. The cells were regularly passaged by detaching with trypsin/EDTA solution, followed by resuspension in DMEM with 10% FBS and seeding into cell culture flasks.

4.4.3 Transfections and incorporation of non-canonical amino acids

HEK293 cells were seeded onto poly-l-lysine (Sigma) coated 24-well plates day prior to transfection. The media was exchanged for Optimem media (Gibco) and transfection mixture of 1 µg DNA and 4 µl of 1 mg.ml polyethyleneimine MW 40,000 (PEI; Polysciences) in Optimem was added to each well. After 4 hours, the media was exchanged for DMEM (supplemented with 10% FBS) with or without Nε–((*tert*-butoxy)carbonyl)-L-lysine (BocK; Bachem) or 3-methyl-L-histidine (Me-His; Bachem).

4.4.4 Flow cytometry

24 hours after transfection, the cells were detached using trypsin/EDTA solution and analysed on Becton Dickinson LSRII SORP flow cytometer (BD Biosciences) with the appropriate filter settings (488 nm coherent sapphire laser for GFP excitation and 561 nm compass laser for mCherry excitation). The front scatter and side scatter was used to identify intact cells and

mean background fluorescence from untransfected cells was subtracted from the measured signal. FlowJo software (FlowJo) was used to analyse the data.

4.4.5 Fluorescence microscopy

The cells were imaged after 24 hours post transfection using Leica DM IL LED inverted fluorescence microscope with QCapture Suite PLUS software.

4.4.6 LC-MS analysis

HEK293 cells were seeded into 75 ml cell culture flasks. After reaching ~ 80% confluency, the media was exchanged for fresh media and the cells were transfected using transfection mixture of 20 μg DNA and 80 μl of 1 mg/ml PEI solution. BocK was added to final concentration of 1 mM and the cells were cultured for 48 hours. The cells were harvested using PBS and lysed in ice cold RIPA buffer supplemented with HALT protease and phosphatase inhibitor cocktail (Sigma). The cells were spun down at 4°C at 16,000 g for 20 minutes, collecting the supernatant. The GFP was purified from the supernatant using GFP-Trap_MA beads (ChromoTek), following the manufacturer's instructions. The GFP was eluted from the beads in 200 mM glycine (pH 2.5) and immediately neutralized by adding 1M Tris base (pH 10.4). The purified GFP was analysed on Agilent 1200 LC-MS with Quadrupole spectrometer (Agilent) using 0.2% formic acid in water and 0.2% formic acid in acetonitrile as buffer A and B, respectively. The samples were run through Phenomenex Jupiter C4 column (150 x 2 mm, 5 μm) into the mass spectrometer. The data was collected in positive mode. The spectra was analysed and deconvoluted in OpenLAB CDS Chemstation software (Agilent). The expected mass was calculated using online tool (http://web.expasy.org/protparam) and corrected for molecular weight of BocK, N-terminal methinone cleavage and N-terminal acetylation.

Chapter 5 Discussion

This thesis presents several advances in the field of genetic code expansion in mammalian cells.

Mammalian cells present an attractive model organism for genetic code expansion for several reasons. Firstly, biological insights gained from studies in human cell lines are of great relevance to understanding of human physiology and pathology. In this regard, genetic code expansion is a unique tool, which allows biologists to study of processes inside living cells with precision on the single amino acid level. Secondly, production of certain biological therapeutics, such as antibody-drug conjugates via genetic code expansion (Oller-Salvia et al., 2018), can only be carried out in eukaryotic cells. Overall, the capabilities of this technology depend both on its efficiency (i.e. how well and how specifically are ncAA incorporated into target proteins) and the variety of ncAA that can be efficiently incorporated.

This thesis addresses both improvements of efficiency in incorporation of ncAAs and the incorporation of new ncAA. Overall, it reports two new orthogonal aaRS-tRNA pairs in mammalian cells (SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ and *Ma*PylRS/tRNA$^{Pyl}(6)_{CUA}$) and the incorporation of two new ncAAs: phosphoserine and its phosphonate analogue. It further provides a method to significantly decrease off-target incorporation of ncAAs at endogenous amber codon sites, further improving the specificity of the method.

In Chapter 2, we demonstrate the incorporation of phosphoserine via the SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ pair and then increase its efficiency via an engineered version of the eukaryotic elongation factor alpha – EF-1α-Sep. We subsequently demonstrate incorporation of non-hydrolysable analague of phosphoserine and use it to synthetically activate a protein kinase Mek1. We note that the approach of using an evolved EF-1α variant could be extended to other ncAA in mammalian cells. The currently used the version of EF-Tu evolved in bacteria, could be further extended using directed evolution in mammalian cells may yield even better variants (Kelemen et al., 2018). Further, the SepRS$^{v1.0}$/tRNA$^{v1.0}_{CUA}$ pair can be evolved to

incorporate amino acids other than phosphoserine (Zhang et al., 2017) and therefore paves the way for genetic encoding of other phosphorylated and non-phosphorylated amino acids in mammalian cells.

In Chater 3, we develop an approach to minimise the off-target incorporation of ncAA in response to the endogenous amber stop codons. We use this approach to improve the signal-to-noise ratio of protein labelling via bio-orthogonal chemistry and subsequently image a microprotein at low expression levels. This approach can be adapted in any applications where the incorporation at endogenous amber codons may be detrimental to the cell health or may interfere with the observed phenomena.

Lastly, in Chapter 4, we report a new orthogonal pair in mammalian cells, the *Ma*PylRS/tRNA$^{Pyl}(6)_{CUA}$. The *Ma*PylRS lacks N-terminal domain found in the majority of PylRS variants and therefore has the potential to be a more versatile synthetase compared to the commonly used PylRS from *M mazei* or *M. barkeri*. We use this pair for simultaneous incorporation of multiple distinct amino acids into a single reporter protein. When paired with appropriate mutually orthogonal chemistries, this approach could be used for multi-probe labelling of proteins inside live cells for imaging and proteomic studies or used in protein cross-linking, leveraging the approaches developed in Chapter 2.

In aggregate, this thesis presents distinct advances in the scope (through two, new orthogonal aaRS/tRNA pairs and two newly encoded ncAA) and efficiency (through minimalisation of off target ncAA incorporation) of genetic code expansion in mammalian cells and should serve as a basis for further improvements in and applications of this technology.

References

Abbe, E. (1873). Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. Archiv für mikroskopische Anatomie 9, 413-418.

Accetto, T., and Avgustin, G. (2011). Inability of Prevotella bryantii to form a functional Shine-Dalgarno interaction reflects unique evolution of ribosome binding sites in Bacteroidetes. PLoS One 6, e22914.

Achenbach, J., and Nierhaus, K.H. (2015). The mechanics of ribosomal translocation. Biochimie 114, 80-89.

Achilli, C., Ciana, A., and Minetti, G. (2015). The discovery of methionine sulfoxide reductase enzymes: An historical account and future perspectives. Biofactors 41, 135-152.

Aerni, H.R., Shifman, M.A., Rogulina, S., O'Donoghue, P., and Rinehart, J. (2015). Revealing the amino acid composition of proteins within an expanded genetic code. Nucleic Acids Res 43, e8.

Ai, H.-w., Lee, J.W., and Schultz, P.G. (2010). A method to site-specifically introduce methyllysine into proteins in E. coli. Chemical communications 46, 5506-5508.

Alff-Steinberger, C., and Epstein, R. (1994). Codon preference in the terminal region of E. coli genes and evolution of stop codon usage. Journal of theoretical biology 168, 461-463.

Allfrey, V., Faulkner, R., and Mirsky, A. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proceedings of the National Academy of Sciences 51, 786-794.

Allis, C.D., Berger, S.L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., and Shiekhattar, R. (2007). New nomenclature for chromatin-modifying enzymes. Cell 131, 633-636.

Allmang, C., and Krol, A. (2006). Selenoprotein synthesis: UGA does not end the story. Biochimie 88, 1561-1571.

Aloush, N., Schvartz, T., König, A., Cohen, S., Tam, B., Akabayov, B., Nachmias, D., Ben-David, O., Elia, N., and Arbely, E. (2018). Reducing Pyrrolysine tRNA Copy Number Improves Live Cell Imaging of Bioorthogonally Labeled Proteins. bioRxiv, 161984.

Ambrogelly, A., Palioura, S., and Söll, D. (2007). Natural expansion of the genetic code. Nature chemical biology 3, 29-35.

Andersen, G.R., Pedersen, L., Valente, L., Chatterjee, I., Kinzy, T.G., Kjeldgaard, M., and Nyborg, J. (2000). Structural basis for nucleotide exchange and competition with tRNA in the yeast elongation factor complex eEF1A: eEF1Bα. Molecular cell 6, 1261-1266.

Ando, R., Hama, H., Yamamoto-Hino, M., Mizuno, H., and Miyawaki, A. (2002). An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. Proceedings of the National Academy of Sciences 99, 12651-12656.

Andrews, S.J., and Rothnagel, J.A. (2014). Emerging evidence for functional peptides encoded by short open reading frames. Nature Reviews Genetics 15, 193.

Arbely, E., Natan, E., Brandt, T., Allen, M.D., Veprintsev, D.B., Robinson, C.V., Chin, J.W., Joerger, A.C., and Fersht, A.R. (2011). Acetylation of lysine 120 of p53 endows DNA-binding specificity at effective physiological salt concentration. Proceedings of the National Academy of Sciences 108, 8251-8256.

Arbely, E., Torres-Kolbus, J., Deiters, A., and Chin, J.W. (2012). Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. Journal of the American Chemical Society 134, 11912-11915.

Arnez, J.G., and Moras, D. (1997). Structural and functional considerations of the aminoacylation reaction. Trends in biochemical sciences 22, 211-216.

Arribere, J.A., Cenik, E.S., Jain, N., Hess, G.T., Lee, C.H., Bassik, M.C., and Fire, A.Z. (2016). Translation readthrough mitigation. Nature 534, 719.

Bae, J.H., Alefelder, S., Kaiser, J.T., Friedrich, R., Moroder, L., Huber, R., and Budisa, N. (2001). Incorporation of β-selenolo [3, 2-b] pyrrolyl-alanine into proteins for phase determination in protein X-ray crystallography1. Journal of molecular biology 309, 925-936.

Bain, J., Diala, E.S., Glabe, C.G., Dix, T.A., and Chamberlin, A.R. (1989). Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. Journal of the American Chemical Society 111, 8013-8014.

Baldini, G., Martoglio, B., Schachenmann, A., Zugliani, C., and Brunner, J. (1988). Mischarging Escherichia coli tRNAPhe with L-4'-[3-(trifluoromethyl)-3H-diazirin-3-yl] phenylalanine, a photoactivatable analog of phenylalanine. Biochemistry 27, 7951-7959.

Baron, C., Heider, J., and Böck, A. (1993). Interaction of translation factor SELB with the formate dehydrogenase H selenopolypeptide mRNA. Proceedings of the National Academy of Sciences 90, 4181-4185.

Baumdick, M., Brüggemann, Y., Schmick, M., Xouri, G., Sabet, O., Davis, L., Chin, J.W., and Bastiaens, P.I. (2015). EGF-dependent re-routing of vesicular recycling switches spontaneous phosphorylation suppression to EGFR signaling. Elife 4, e12223.

Beranek, V., Reinkemeier, C.D., Zhang, M.S., Liang, A.D., Kym, G., and Chin, J.W. (2018). Genetically Encoded Protein Phosphorylation in Mammalian Cells. Cell Chem Biol 25, 1–8.

Berry, M.J., Banu, L., Chen, Y., Mandel, S.J., Kieffer, J.D., Harney, J.W., and Larsen, P.R. (1991). Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3′ untranslated region. Nature 353, 273.

Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J., and Hess, H.F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. Science 313, 1642-1645.

Bianco, A., Townsley, F.M., Greiss, S., Lang, K., and Chin, J.W. (2012). Expanding the genetic code of Drosophila melanogaster. Nature chemical biology 8, 748.

Björk, G.R. (1995). Genetic Dissection of Synthesis and Function of Modified Nucleosides in Bacterial Transfer RNA1. In Progress in nucleic acid research and molecular biology (Elsevier), pp. 263-338.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., and Mayhew, G.F. (1997). The complete genome sequence of Escherichia coli K-12. science 277, 1453-1462.

Blight, S.K., Larue, R.C., Mahapatra, A., Longstaff, D.G., Chang, E., Zhao, G., Kang, P.T., Green-Church, K.B., Chan, M.K., and Krzycki, J.A. (2004). Direct charging of tRNA CUA with pyrrolysine in vitro and in vivo. Nature 431, 333.

Blume-Jensen, P., and Hunter, T. (2001). Oncogenic kinase signalling. Nature 411, 355.

Boersema, P.J., Foong, L.Y., Ding, V.M., Lemeer, S., van Breukelen, B., Philp, R., Boekhorst, J., Snel, B., den Hertog, J., and Choo, A.B. (2010). In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labelling. Molecular & Cellular Proteomics 9, 84-99.

Boger, D.L. (1986). Diels-Alder reactions of heterocyclic aza dienes. Scope and applications. Chemical Reviews 86, 781-793.

Borrel, G., Gaci, N., Peyret, P., O'Toole, P.W., Gribaldo, S., and Brugere, J.F. (2014). Unique characteristics of the pyrrolysine system in the 7th order of methanogens: implications for the evolution of a genetic code expansion cassette. Archaea 2014, 374146.

Borrel, G., O'Toole, P.W., Harris, H.M., Peyret, P., Brugere, J.-F., and Gribaldo, S. (2013). Phylogenomic data support a seventh order of methylotrophic methanogens and provide insights into the evolution of methanogenesis. Genome biology and evolution 5, 1769-1780.

Borrmann, A., Milles, S., Plass, T., Dommerholt, J., Verkade, J.M., Wießler, M., Schultz, C., van Hest, J.C., van Delft, F.L., and Lemke, E.A. (2012). Genetic Encoding of a Bicyclo [6.1. 0] nonyne-Charged Amino Acid Enables Fast Cellular Protein Imaging by Metal-Free Ligation. ChemBioChem 13, 2094-2099.

Brick, P., Bhat, T., and Blow, D. (1989). Structure of tyrosyl-tRNA synthetase refined at 2.3 Å resolution: Interaction of the enzyme with the tyrosyl adenylate intermediate. Journal of molecular biology 208, 83-98.

Brown, A., Shao, S., Murray, J., Hegde, R.S., and Ramakrishnan, V. (2015). Structural basis for stop codon recognition in eukaryotes. Nature 524, 493.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell 84, 843-851.

Budisa, N., Alefelder, S., Bae, J.H., Golbik, R., Minks, C., Huber, R., and Moroder, L. (2001). Proteins with β-(thienopyrrolyl) alanines as alternative chromophores and pharmaceutically active amino acids. Protein Science 10, 1281-1292.

Budisa, N., Minks, C., Alefelder, S., Wenger, W., Dong, F., Moroder, L., and Huber, R. (1999). Toward the experimental codon reassignment in vivo: protein building with an expanded amino acid repertoire. The FASEB journal 13, 41-51.

Capecchi, M.R. (1967). Polypeptide chain termination in vitro: isolation of a release factor. Proceedings of the National Academy of Sciences 58, 1144-1151.

Carvalho, M.D.G.D.C., Carvalho, J.F., and Merrick, W.C. (1984). Biological characterization of various forms of elongation factor 1 from rabbit reticulocytes. Archives of Biochemistry and Biophysics 234, 603-611.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science 263, 802-805.

Chambers, I., Frampton, J., Goldfarb, P., Affara, N., McBain, W., and Harrison, P.R. (1986). The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the ‘termination’codon, TGA. The EMBO journal 5, 1221-1227.

Chang, B., Halgamuge, S., and Tang, S.-L. (2006). Analysis of SD sequences in completed microbial genomes: non-SD-led genes are as common as SD-led genes. Gene 373, 90-99.

Chang, Y.-F., Imam, J.S., and Wilkinson, M.F. (2007). The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem 76, 51-74.

Chapeville, F., Lipmann, F., Von Ehrenstein, G., Weisblum, B., Ray, W.J., and Benzer, S. (1962). On the role of soluble ribonucleic acid in coding for amino acids. Proceedings of the National Academy of Sciences 48, 1086-1092.

Chehab, N.H., Malikzay, A., Stavridi, E.S., and Halazonetis, T.D. (1999). Phosphorylation of Ser-20 mediates stabilization of human p53 in response to DNA damage. Proceedings of the National Academy of Sciences 96, 13777-13782.

Chen, H.-T., Warfield, L., and Hahn, S. (2007). The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. Nature Structural and Molecular Biology 14, 696.

Chen, P.R., Groff, D., Guo, J., Ou, W., Cellitti, S., Geierstanger, B.H., and Schultz, P.G. (2009). A facile system for encoding unnatural amino acids in mammalian cells. Angewandte Chemie 121, 4112-4115.

Chin, J.W. (2014). Expanding and reprogramming the genetic code of cells and animals. Annu Rev Biochem 83, 379-408.

Chin, J.W. (2017). Expanding and reprogramming the genetic code. Nature 550, 53-60.

Chin, J.W., Cropp, T.A., Anderson, J.C., Mukherji, M., Zhang, Z., and Schultz, P.G. (2003). An expanded eukaryotic genetic code. Science 301, 964-967.

Chin, J.W., Martin, A.B., King, D.S., Wang, L., and Schultz, P.G. (2002a). Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. Proceedings of the National Academy of Sciences 99, 11020-11024.

Chin, J.W., Santoro, S.W., Martin, A.B., King, D.S., Wang, L., and Schultz, P.G. (2002b). Addition of p-Azido-l-phenylalanine to the Genetic Code of Escherichia c oli. Journal of the American Chemical Society 124, 9026-9027.

Chou, C., Uprety, R., Davis, L., Chin, J.W., and Deiters, A. (2011). Genetically encoding an aliphatic diazirine for protein photocrosslinking. Chemical Science 2, 480-483.

Cohen, P. (2002). The origins of protein phosphorylation. Nature cell biology 4, E127.

Cone, J.E., Del Rio, R.M., Davis, J.N., and Stadtman, T.C. (1976). Chemical characterization of the selenoprotein component of clostridial glycine reductase: identification of selenocysteine as the organoselenium moiety. Proceedings of the National Academy of Sciences 73, 2659-2663.

Coons, A.H. (1942). The demonstration of pneumococcal antigen in tissues by the use of fluorescent antibody. J Immunol 45, 159.

Copeland, P.R., Fletcher, J.E., Carlson, B.A., Hatfield, D.L., and Driscoll, D.M. (2000). A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. The EMBO journal 19, 306-314.

Cornish, V.W., Mendel, D., and Schultz, P.G. (1995). Probing protein structure and function with an expanded genetic code. Angewandte Chemie International Edition in English 34, 621-633.

Cowie, D.B., and Cohen, G.N. (1957). Biosynthesis by Escherichia coli of active altered proteins containing selenium instead of sulfur. Biochimica et biophysica acta 26, 252-261.

Crepin, T., Shalak, V.F., Yaremchuk, A.D., Vlasenko, D.O., McCarthy, A., Negrutskii, B.S., Tukalo, M.A., and El'skaya, A.V. (2014). Mammalian translation elongation factor eEF1A2: X-ray structure and new features of GDP/GTP exchange mechanism in higher eukaryotes. Nucleic acids research 42, 12939-12948.

Crick, F., Barnett, L., Brenner, S., and Watts-Tobin, R.J. (1961). General nature of the genetic code for proteins.

Crick, F.H. (1968). The origin of the genetic code. Journal of molecular biology 38, 367-379.

D'Lima, N.G., Ma, J., Winkler, L., Chu, Q., Loh, K.H., Corpuz, E.O., Budnik, B.A., Lykke-Andersen, J., Saghatelian, A., and Slavoff, S.A. (2017). A human microprotein that interacts with the mRNA decapping complex. Nature chemical biology 13, 174.

Datta, D., Wang, P., Carrico, I.S., Mayo, S.L., and Tirrell, D.A. (2002). A designed phenylalanyl-tRNA synthetase variant allows efficient in vivo incorporation of aryl ketone functionality into proteins. Journal of the American Chemical Society 124, 5652-5653.

Dawson, P.E., Muir, T.W., Clark-Lewis, I., and Kent, S. (1994). Synthesis of proteins by native chemical ligation. Science 266, 776-779.

Deiters, A., Cropp, T.A., Mukherji, M., Chin, J.W., Anderson, J.C., and Schultz, P.G. (2003). Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae. Journal of the American Chemical Society 125, 11782-11783.

Dempsey, G.T., Vaughan, J.C., Chen, K.H., Bates, M., and Zhuang, X. (2011). Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. Nature methods 8, 1027.

Deribe, Y.L., Pawson, T., and Dikic, I. (2010). Post-translational modifications in signal integration. Nature structural & molecular biology 17, 666.

Di Cerbo, V., Mohn, F., Ryan, D.P., Montellier, E., Kacem, S., Tropberger, P., Kallis, E., Holzner, M., Hoerner, L., and Feldmann, A. (2014). Acetylation of histone H3 at lysine 64 regulates nucleosome dynamics and facilitates transcription. elife 3, e01632.

Doll, S., and Burlingame, A.L. (2014). Mass spectrometry-based detection and assignment of protein posttranslational modifications. ACS chemical biology 10, 63-71.

Dougherty, D.A., and Van Arnam, E.B. (2014). In Vivo Incorporation of Non-canonical Amino Acids by Using the Chemical Aminoacylation Strategy: A Broadly Applicable Mechanistic Tool. ChemBioChem 15, 1710-1720.

Drazic, A., Myklebust, L.M., Ree, R., and Arnesen, T. (2016). The world of protein acetylation. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics 1864, 1372-1401.

Dumas, A., Lercher, L., Spicer, C.D., and Davis, B.G. (2015). Designing logical codon reassignment–Expanding the chemistry in biology. Chemical science 6, 50-69.

Durek, T., and Becker, C.F. (2005). Protein semi-synthesis: new proteins for functional and structural studies. Biomolecular engineering 22, 153-172.

Elangovan, M., Day, R., and Periasamy, A. (2002). Nanosecond fluorescence resonance energy transfer-fluorescence lifetime imaging microscopy to localize the protein interactions in a single living cell. Journal of microscopy 205, 3-14.

Elliott, T.S., Bianco, A., Townsley, F.M., Fried, S.D., and Chin, J.W. (2016). Tagging and Enriching Proteins Enables Cell-Specific Proteomics. Cell Chem Biol 23, 805-815.

Elliott, T.S., Townsley, F.M., Bianco, A., Ernst, R.J., Sachdeva, A., Elsässer, S.J., Davis, L., Lang, K., Pisa, R., and Greiss, S. (2014). Proteome labelling and protein identification in specific tissues and at specific developmental stages in an animal. Nat Biotechnol 32, 465.

Elsässer, S.J., Ernst, R.J., Walker, O.S., and Chin, J.W. (2016). Genetic code expansion in stable cell lines enables encoded chromatin modification. Nature methods 13, 158.

Fagegaltier, D., Hubert, N., Yamada, K., Mizutani, T., Carbon, P., and Krol, A. (2000). Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. The EMBO journal 19, 4796-4805.

Fan, C., and Bobik, T.A. (2008). The PduX enzyme of Salmonella enterica is an L-threonine kinase used for coenzyme B12 synthesis. Journal of Biological Chemistry 283, 11322-11329.

Fan, C., Fromm, H.J., and Bobik, T.A. (2009). Kinetic and functional analysis of L-threonine kinase, the PduX enzyme of Salmonella enterica. Journal of Biological Chemistry, jbc. M109. 027425.

Fan, C., Ip, K., and Söll, D. (2016). Expanding the genetic code of Escherichia coli with phosphotyrosine. FEBS letters 590, 3040-3047.

Fechter, P., Rudinger-Thirion, J., Tukalo, M., and Giegé, R. (2001). Major tyrosine identity determinants in Methanococcus jannaschii and Saccharomyces cerevisiae tRNATyr are conserved but expressed differently. European journal of biochemistry 268, 761-767.

Fekner, T., Li, X., Lee, M.M., and Chan, M.K. (2009). A pyrrolysine analogue for protein click chemistry. Angewandte Chemie 121, 1661-1663.

Fersht, A.R. (1977). Editing mechanisms in protein synthesis. Rejection of valine by the isoleucyl-tRNA synthetase. Biochemistry 16, 1025-1030.

Fersht, A.R., and Dingwall, C. (1979). Cysteinyl-tRNA synthetase from Escherichia coli does not need an editing mechanism to reject serine and alanine. High binding energy of small groups in specific molecular interactions. Biochemistry 18, 1245-1249.

Fersht, A.R., Shindler, J.S., and Tsui, W.-C. (1980). Probing the limits of protein-amino acid side chain recognition with the aminoacyl-tRNA synthetases. Discrimination against phenylalanine by tyrosyl-tRNA synthetases. Biochemistry 19, 5520-5524.

Fletcher, J.E., Copeland, P.R., Driscoll, D.M., and Krol, A. (2001). The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2. Rna 7, 1442-1453.

Forchhammer, K., Leinfelder, W., and Böck, A. (1989). Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. Nature 342, 453.

Frolova, L., Le Goff, X., Rasmussen, H.H., Cheperegin, S., Drugeon, G., Kress, M., Arman, I., Haenni, A.-L., Celis, J.E., and Phllippe, M. (1994). A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. Nature 372, 701.

Fukunaga, R., and Yokoyama, S. (2007). Structural insights into the first step of RNA-dependent cysteine biosynthesis in archaea. Nature Structural and Molecular Biology 14, 272.

Garza, D., Medhora, M., and Hartl, D. (1990). Drosophila nonsense suppressors: functional analysis in Saccharomyces cerevisiae, Drosophila tissue culture cells and Drosophila melanogaster. Genetics 126, 625-637.

Gaston, M.A., Zhang, L., Green-Church, K.B., and Krzycki, J.A. (2011). The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. Nature 471, 647.

Gautier, A., Deiters, A., and Chin, J.W. (2011). Light-activated kinases enable temporal dissection of signaling networks in living cells. Journal of the American Chemical Society 133, 2124-2127.

Gautier, A., Juillerat, A., Heinis, C., Corrêa Jr, I.R., Kindermann, M., Beaufils, F., and Johnsson, K. (2008). An engineered protein tag for multiprotein labelling in living cells. Chemistry & biology 15, 128-136.

Gautier, A., Nguyen, D.P., Lusic, H., An, W., Deiters, A., and Chin, J.W. (2010). Genetically encoded photocontrol of protein localization in mammalian cells. Journal of the American Chemical Society 132, 4086-4088.

Ge, Y., Fan, X., and Chen, P.R. (2016). A genetically encoded multifunctional unnatural amino acid for versatile protein manipulations in living cells. Chemical science 7, 7055-7060.

Grangeasse, C., Stülke, J., and Mijakovic, I. (2015). Regulatory potential of post-translational modifications in bacteria. Frontiers in microbiology 6, 500.

Greiss, S., and Chin, J.W. (2011). Expanding the genetic code of an animal. Journal of the American Chemical Society 133, 14196-14199.

Grimm, J.B., English, B.P., Chen, J., Slaughter, J.P., Zhang, Z., Revyakin, A., Patel, R., Macklin, J.J., Normanno, D., and Singer, R.H. (2015). A general method to improve fluorophores for live-cell and single-molecule microscopy. Nature methods 12, 244.

Grimm, J.B., Muthusamy, A.K., Liang, Y., Brown, T.A., Lemon, W.C., Patel, R., Lu, R., Macklin, J.J., Keller, P.J., and Ji, N. (2017). A general method to fine-tune fluorophores for live-cell and in vivo imaging. Nature methods 14, 987.

Groff, D., Chen, P.R., Peters, F.B., and Schultz, P.G. (2010). A Genetically Encoded ε-N-Methyl Lysine in Mammalian Cells. ChemBioChem 11, 1066-1068.

Gualerzi, C.O., and Pon, C.L. (2015). Initiation of mRNA translation in bacteria: structural and dynamic aspects. Cellular and Molecular Life Sciences 72, 4341-4367.

Gustafsson, M.G. (2005). Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution. Proceedings of the National Academy of Sciences 102, 13081-13086.

Hamano-Takaku, F., Iwama, T., Saito-Yano, S., Takaku, K., Monden, Y., Kitabatake, M., Söll, D., and Nishimura, S. (2000). A mutant Escherichia coli tyrosyl-tRNA synthetase utilizes the unnatural amino acid azatyrosine more efficiently than tyrosine. Journal of Biological Chemistry 275, 40324-40328.

Hancock, S.M., Uprety, R., Deiters, A., and Chin, J.W. (2010). Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. Journal of the American Chemical Society 132, 14819-14824.

Hao, B., Gong, W., Ferguson, T.K., James, C.M., Krzycki, J.A., and Chan, M.K. (2002). A new UAG-encoded residue in the structure of a methanogen methyltransferase. Science 296, 1462-1466.

Hartz, D., Binkley, J., Hollingsworth, T., and Gold, L. (1990). Domains of initiator tRNA and initiation codon crucial for initiator tRNA selection by Escherichia coli IF3. Genes & development 4, 1790-1800.

Hatfield, D., Choi, I.S., Mischke, S., and Owens, L. (1992). Selenocysteyl-tRNAs recognize UGA in Betavulgaris, a higher plant, and in Gliocladium virens, a filamentous fungus. Biochemical and biophysical research communications 184, 254-259.

Hecht, S.M., Alford, B., Kuroda, Y., and Kitano, S. (1978). " Chemical aminoacylation" of tRNA's. Journal of Biological Chemistry 253, 4517-4520.

Heckler, T.G., Chang, L.H., Zama, Y., Naka, T., Chorghade, M.S., and Hecht, S.M. (1984). T4 RNA ligase mediated preparation of novel" chemically misacylated" tRNAPheS. Biochemistry 23, 1468-1473.

Heim, R., and Tsien, R.Y. (1996). Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. Current biology 6, 178-182.

Heinemann, I.U., O'Donoghue, P., Madinger, C., Benner, J., Randau, L., Noren, C.J., and Söll, D. (2009). The appearance of pyrrolysine in tRNAHis guanylyltransferase by neutral evolution. Proceedings of the National Academy of Sciences 106, 21103-21108.

Hell, S.W., and Wichmann, J. (1994). Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. Optics letters 19, 780-782.

Hemphill, J., Chou, C., Chin, J.W., and Deiters, A. (2013). Genetically encoded light-activated transcription for spatiotemporal control of gene expression and gene silencing in mammalian cells. Journal of the American Chemical Society 135, 13433-13439.

Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science 254, 51-58.

Herring, S., Ambrogelly, A., Gundllapalli, S., O'Donoghue, P., Polycarpo, C.R., and Söll, D. (2007). The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity. FEBS Lett 581, 3197-3203.

Hess, S.T., Girirajan, T.P., and Mason, M.D. (2006). Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. Biophysical journal 91, 4258-4272.

Himo, F., Lovell, T., Hilgraf, R., Rostovtsev, V.V., Noodleman, L., Sharpless, K.B., and Fokin, V.V. (2005). Copper (I)-catalyzed synthesis of azoles. DFT study predicts unprecedented reactivity and intermediates. Journal of the American Chemical Society 127, 210-216.

Hodgkin, J. (1985). Novel nematode amber suppressors. Genetics 111, 287-310.

Hodgkin, J., Papp, A., Pulak, R., Ambros, V., and Anderson, P. (1989). A new kind of informational suppression in the nematode Caenorhabditis elegans. Genetics 123, 301-313.

Hoffmann, J.E., Dziuba, D., Stein, F., and Schultz, C. (2018). A Bifunctional Noncanonical Amino Acid: Synthesis, Expression, and Residue-Specific Proteome-wide Incorporation. Biochemistry 57, 4747-4752.

Hohn, M.J., Park, H.-S., O'Donoghue, P., Schnitzbauer, M., and Söll, D. (2006). Emergence of the universal genetic code imprinted in an RNA record. Proceedings of the National Academy of Sciences 103, 18095-18100.

Hohsaka, T., Ashizuka, Y., Murakami, H., and Sisido, M. (1996). Incorporation of nonnatural amino acids into streptavidin through in vitro frame-shift suppression. Journal of the American Chemical Society 118, 9778-9779.

Hondal, R.J., Marino, S.M., and Gladyshev, V.N. (2013). Selenocysteine in thiol/disulfide-like exchange reactions. Antioxidants & redox signaling 18, 1675-1689.

Hoppmann, C., Wong, A., Yang, B., Li, S., Hunter, T., Shokat, K.M., and Wang, L. (2017). Site-specific incorporation of phosphotyrosine using an expanded genetic code. Nature chemical biology 13, 842.

Hori, H. (2014). Methylated nucleosides in tRNA and tRNA methyltransferases. Frontiers in genetics 5, 144.

Hsu, P.P., Kang, S.A., Rameseder, J., Zhang, Y., Ottina, K.A., Lim, D., Peterson, T.R., Choi, Y., Gray, N.S., and Yaffe, M.B. (2011). The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. Science 332, 1317-1322.

Huang, B., Bates, M., and Zhuang, X. (2009). Super-resolution fluorescence microscopy. Annual review of biochemistry 78, 993-1016.

Huang, Y., Wan, W., Russell, W.K., Pai, P.-J., Wang, Z., Russell, D.H., and Liu, W. (2010). Genetic incorporation of an aliphatic keto-containing amino acid into proteins for their site-specific modifications. Bioorganic & medicinal chemistry letters 20, 878-880.

Hughes, R.A., and Ellington, A.D. (2010). Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA. Nucleic Acids Res 38, 6813-6830.

Huguenin-Dezot, N., De Cesare, V., Peltier, J., Knebel, A., Kristaryianto, Y.A., Rogerson, D.T., Kulathu, Y., Trost, M., and Chin, J.W. (2016). Synthesis of isomeric phosphoubiquitin chains reveals that phosphorylation controls deubiquitinase activity and specificity. Cell reports 16, 1180-1193.

Humphrey, S.J., Azimifar, S.B., and Mann, M. (2015). High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. Nature biotechnology 33, 990.

Humphrey, S.J., Yang, G., Yang, P., Fazakerley, D.J., Stöckli, J., Yang, J.Y., and James, D.E. (2013). Dynamic adipocyte phosphoproteome reveals that Akt directly regulates mTORC2. Cell metabolism 17, 1009-1020.

Ibba, M., Kast, P., and Hennecke, H. (1994). Substrate specificity is determined by amino acid binding pocket size in Escherichia coli phenylalanyl-tRNA synthetase. Biochemistry 33, 7107-7112.

Ibba, M., and Söll, D. (2000). Aminoacyl-tRNA synthesis. Annual review of biochemistry 69, 617-650.

Iraha, F., Oki, K., Kobayashi, T., Ohno, S., Yokogawa, T., Nishikawa, K., Yokoyama, S., and Sakamoto, K. (2010). Functional replacement of the endogenous tyrosyl-tRNA synthetase-tRNATyr pair by the archaeal tyrosine pair in Escherichia coli for genetic code expansion. Nucleic Acids Res 38, 3682-3691.

Isaacs, F.J., Carr, P.A., Wang, H.H., Lajoie, M.J., Sterling, B., Kraal, L., Tolonen, A.C., Gianoulis, T.A., Goodman, D.B., and Reppas, N.B. (2011). Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science 333, 348-353.

Italia, J.S., Addy, P.S., Wrobel, C.J., Crawford, L.A., Lajoie, M.J., Zheng, Y., and Chatterjee, A. (2017). An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. Nat Chem Biol 13, 446.

Italia, J.S., Latour, C., Wrobel, C.J., and Chatterjee, A. (2018). Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes. Cell Chem Biol.

Ito, K., Ebihara, K., Uno, M., and Nakamura, Y. (1996). Conserved motifs in prokaryotic and eukaryotic polypeptide release factors: tRNA-protein mimicry hypothesis. Proceedings of the National Academy of Sciences 93, 5443-5448.

Jakubowski, H., and Goldman, E. (1992). Editing of errors in selection of amino acids for protein synthesis. Microbiological reviews 56, 412-429.

Jans, D.A., and Hubner, S. (1996). Regulation of protein transport to the nucleus: central role of phosphorylation. Physiological reviews 76, 651-685.

Jiang, R., and Krzycki, J.A. (2012). PylSn and the homologous N-terminal domain of pyrrolysyl-tRNA synthetase bind the tRNA that is essential for the genetic encoding of pyrrolysine. J Biol Chem, jbc. M112. 396754.

Johnson, D.B., Xu, J., Shen, Z., Takimoto, J.K., Schultz, M.D., Schmitz, R.J., Xiang, Z., Ecker, J.R., Briggs, S.P., and Wang, L. (2011). RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. Nature chemical biology 7, 779.

Kast, P., and Hennecke, H. (1991). Amino acid substrate specificity of Escherichia coli phenylalanyl-tRNA synthetase altered by distinct mutations. Journal of molecular biology 222, 99-124.

Kavran, J.M., Gundllapalli, S., O'Donoghue, P., Englert, M., Söll, D., and Steitz, T.A. (2007). Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. Proceedings of the National Academy of Sciences 104, 11268-11273.

Kaya, E., Gutsmiedl, K., Vrabel, M., Müller, M., Thumbs, P., and Carell, T. (2009). Synthesis of threefold glycosylated proteins using click chemistry and genetically encoded unnatural amino acids. ChemBioChem 10, 2858-2861.

Kedersha, N., Stoecklin, G., Ayodele, M., Yacono, P., Lykke-Andersen, J., Fritzler, M.J., Scheuner, D., Kaufman, R.J., Golan, D.E., and Anderson, P. (2005). Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. J Cell Biol 169, 871-884.

Kelemen, R. E., Erickson, S. B., & Chatterjee, A. (2018). Synthesis at the interface of virology and genetic code expansion. Current opinion in chemical biology, 46, 164-171.

Keller, P.J. (2013). Imaging morphogenesis: technological advances and biological insights. Science 340, 1234168.

Keller, P.J., Schmidt, A.D., Wittbrodt, J., and Stelzer, E.H. (2008). Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. science 322, 1065-1069.

Khorana, H.G., Büuchi, H., Ghosh, H., Gupta, N., Jacob, T., Kössel, H., Morgan, R., Narang, S., Ohtsuka, E., and Wells, R. (1966). Polynucleotide synthesis and the genetic code. Paper presented at: Cold Spring Harbor Symposia on Quantitative Biology (Cold Spring Harbor Laboratory Press).

Kim, C.H., Kang, M., Kim, H.J., Chatterjee, A., and Schultz, P.G. (2012). Site-Specific Incorporation of ε-N-Crotonyllysine into Histones. Angewandte Chemie International Edition 51, 7246-7249.

Kimble, J., and Hirsh, D. (1979). The postembryonic cell lineages of the hermaphrodite and male gonads in Caenorhabditis elegans. Developmental biology 70, 396-417.

Kirshenbaum, K., Carrico, I.S., and Tirrell, D.A. (2002). Biosynthesis of proteins incorporating a versatile set of phenylalanine analogues. ChemBioChem 3, 235-237.

Klar, T.A., Jakobs, S., Dyba, M., Egner, A., and Hell, S.W. (2000). Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. Proceedings of the National Academy of Sciences 97, 8206-8210.

Knight, R.D., Freeland, S.J., and Landweber, L.F. (2001). Rewiring the keyboard: evolvability of the genetic code. Nature Reviews Genetics 2, 49.

Köhrer, C., Xie, L., Kellerer, S., Varshney, U., and RajBhandary, U.L. (2001). Import of amber and ochre suppressor tRNAs into mammalian cells: a general approach to site-specific insertion of amino acid analogues into proteins. Proceedings of the National Academy of Sciences 98, 14310-14315.

Kondo, K., Hodgkin, J., and Waterston, R.H. (1988). Differential expression of five tRNA (UAGTrp) amber suppressors in Caenorhabditis elegans. Molecular and cellular biology 8, 3627-3635.

Kondo, K., Makovec, B., Waterston, R.H., and Hodgkin, J. (1990). Genetic and molecular analysis of eight tRNATrp amber suppressors in Caenorhabditis elegans. Journal of molecular biology 215, 7-19.

Korostelev, A.A. (2011). Structural aspects of translation termination on the ribosome. Rna.

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic acids research 15, 8125-8148.

Krall, N., Da Cruz, F.P., Boutureira, O., and Bernardes, G.J. (2016). Site-selective protein-modification chemistry for basic biology and drug development. Nature chemistry 8, 103.

Kramer, E.B., and Farabaugh, P.J. (2007). The frequency of translational misreading errors in E. coli is largely determined by tRNA competition. Rna 13, 87-96.

Krogager, T.P., Ernst, R.J., Elliott, T.S., Calo, L., Beranek, V., Ciabatti, E., Spillantini, M.G., Tripodi, M., Hastings, M.H., and Chin, J.W. (2018). Labelling and identifying cell-specific proteomes in the mouse brain. Nat Biotechnol 36, 156-159.

Kryukov, G.V., Castellano, S., Novoselov, S.V., Lobanov, A.V., Zehtab, O., Guigó, R., and Gladyshev, V.N. (2003). Characterization of mammalian selenoproteomes. Science 300, 1439-1443.

Kryukov, G.V., and Gladyshev, V.N. (2004). The prokaryotic selenoproteome. EMBO reports 5, 538-543.

Krzycki, J.A. (2004). Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases. Current opinion in chemical biology 8, 484-491.

Lajoie, M.J., Rovner, A.J., Goodman, D.B., Aerni, H.-R., Haimovich, A.D., Kuznetsov, G., Mercer, J.A., Wang, H.H., Carr, P.A., and Mosberg, J.A. (2013). Genomically recoded organisms expand biological functions. science 342, 357-360.

Lammers, M., Neumann, H., Chin, J.W., and James, L.C. (2010). Acetylation regulates cyclophilin A catalysis, immunosuppression and HIV isomerization. Nature chemical biology 6, 331.

Lang, K., and Chin, J.W. (2014). Cellular incorporation of unnatural amino acids and bioorthogonal labelling of proteins. Chemical reviews 114, 4764-4806.

Lang, K., Davis, L., Torres-Kolbus, J., Chou, C., Deiters, A., and Chin, J.W. (2012a). Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. Nature chemistry 4, 298-304.

Lang, K., Davis, L., Wallace, S., Mahesh, M., Cox, D.J., Blackman, M.L., Fox, J.M., and Chin, J.W. (2012b). Genetic encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labelling in vitro and in live mammalian cells via rapid fluorogenic Diels–Alder reactions. Journal of the American Chemical Society 134, 10317-10320.

LaRiviere, F.J., Wolfson, A.D., and Uhlenbeck, O.C. (2001). Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation. Science 294, 165-168.

Laski, F.A., Ganguly, S., Sharp, P.A., RajBhandary, U.L., and Rubin, G.M. (1989). Construction, stable transformation, and function of an amber suppressor tRNA gene in Drosophila melanogaster. Proceedings of the National Academy of Sciences 86, 6696-6698.

Lavis, L.D. (2017). Chemistry is dead. Long live chemistry! Biochemistry 56, 5165-5170.

Lavis, L.D., and Raines, R.T. (2008). Bright ideas for chemical biology. ACS chemical biology 3, 142-155.

Lee, M.M., Fekner, T., Tang, T.H., Wang, L., Chan, A.H.Y., Hsu, P.H., Au, S.W., and Chan, M.K. (2013). A Click-and-Release Pyrrolysine Analogue. ChemBioChem 14, 805-808.

Leinfelder, W., Zehelein, E., MandrandBerthelot, M., and Bock, A. (1988). Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. Nature 331, 723.

Lemeer, S., and Heck, A.J. (2009). The phosphoproteomics data explosion. Current opinion in chemical biology 13, 414-420.

Levy, E.D., Michnick, S.W., and Landry, C.R. (2012). Protein abundance is key to distinguish promiscuous from functional phosphorylation based on evolutionary information. Philosophical Transactions of the Royal Society of London B: Biological Sciences 367, 2594-2606.

Li, W.-T., Mahapatra, A., Longstaff, D.G., Bechtel, J., Zhao, G., Kang, P.T., Chan, M.K., and Krzycki, J.A. (2009a). Specificity of pyrrolysyl-tRNA synthetase for pyrrolysine and pyrrolysine analogs. Journal of molecular biology 385, 1156-1164.

Li, X., Fekner, T., and Chan, M.K. (2010). N6-(2-(R)-Propargylglycyl) lysine as a Clickable Pyrrolysine Mimic. Chemistry–An Asian Journal 5, 1765-1769.

Li, X., Fekner, T., Ottesen, J.J., and Chan, M.K. (2009b). A pyrrolysine analogue for site-specific protein ubiquitination. Angewandte Chemie International Edition 48, 9184-9187.

Liu, C.C., and Schultz, P.G. (2006). Recombinant expression of selectively sulfated proteins in Escherichia coli. Nature biotechnology 24, 1436.

Liu, C.C., and Schultz, P.G. (2010). Adding new chemistries to the genetic code. Annual review of biochemistry 79, 413-444.

Liu, D.R., Magliery, T.J., Pastrnak, M., and Schultz, P.G. (1997a). Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. Proceedings of the National Academy of Sciences 94, 10092-10097.

Liu, D.R., Magliery, T.J., and Schultz, P.G. (1997b). Characterization of an ‘orthogonal’suppressor tRNA derived from E. coli tRNA2Gln. Chemistry & biology 4, 685-691.

Liu, D.R., and Schultz, P.G. (1999). Progress toward the evolution of an organism with an expanded genetic code. Proceedings of the National Academy of Sciences 96, 4780-4785.

Liu, H., Wang, L., Brock, A., Wong, C.-H., and Schultz, P.G. (2003). A method for the generation of glycoprotein mimetics. Journal of the American Chemical Society 125, 1702-1703.

Liu, W., Brock, A., Chen, S., Chen, S., and Schultz, P.G. (2007). Genetic incorporation of unnatural amino acids into proteins in mammalian cells. Nature methods 4, 239.

Liu, X., Long, F., Peng, H., Aerni, S.J., Jiang, M., Sánchez-Blanco, A., Murray, J.I., Preston, E., Mericle, B., and Batzoglou, S. (2009). Analysis of cell fate from single-cell gene expression profiles in C. elegans. Cell 139, 623-633.

Loftfield, R.B., and Vanderjagt, D. (1972). The frequency of errors in protein biosynthesis. Biochemical Journal 128, 1353.

Longman, D., Plasterk, R.H., Johnstone, I.L., and Cáceres, J.F. (2007). Mechanistic insights and identification of two novel factors in the C. elegans NMD pathway. Genes & development 21, 000-000.

Longstaff, D.G., Larue, R.C., Faust, J.E., Mahapatra, A., Zhang, L., Green-Church, K.B., and Krzycki, J.A. (2007). A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. Proceedings of the National Academy of Sciences 104, 1021-1026.

Los, G.V., Encell, L.P., McDougall, M.G., Hartzell, D.D., Karassina, N., Zimprich, C., Wood, M.G., Learish, R., Ohana, R.F., and Urh, M. (2008). HaloTag: a novel protein labelling technology for cell imaging and protein analysis. ACS chemical biology 3, 373-382.

Louie, A., and Jurnak, F. (1985). Kinetic studies of Escherichia coli elongation factor Tu-guanosine 5'-triphosphate-aminoacyl-tRNA complexes. Biochemistry 24, 6433-6439.

Lozupone, C.A., Knight, R.D., and Landweber, L.F. (2001). The molecular basis of nuclear genetic code change in ciliates. Current Biology 11, 65-74.

Lukinavičius, G., Blaukopf, C., Pershagen, E., Schena, A., Reymond, L., Derivery, E., Gonzalez-Gaitan, M., D'Este, E., Hell, S.W., and Gerlich, D.W. (2015). SiR–Hoechst is a far-red DNA stain for live-cell nanoscopy. Nature communications 6, 8497.

Lukinavičius, G., Reymond, L., D'este, E., Masharina, A., Göttfert, F., Ta, H., Güther, A., Fournier, M., Rizzo, S., and Waldmann, H. (2014). Fluorogenic probes for live-cell imaging of the cytoskeleton. Nature methods 11, 731.

Lukinavičius, G., Umezawa, K., Olivier, N., Honigmann, A., Yang, G., Plass, T., Mueller, V., Reymond, L., Corrêa Jr, I.R., and Luo, Z.-G. (2013). A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins. Nature chemistry 5, 132-139.

Lukinavičius, G.v., Reymond, L., Umezawa, K., Sallin, O., D'Este, E., Göttfert, F., Ta, H., Hell, S.W., Urano, Y., and Johnsson, K. (2016). Fluorogenic probes for multicolor imaging in living cells. Journal of the American Chemical Society 138, 9365-9368.

Lundby, A., Secher, A., Lage, K., Nordsborg, N.B., Dmytriyev, A., Lundby, C., and Olsen, J.V. (2012). Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues. Nature communications 3, 876.

Luo, X., Fu, G., Wang, R.E., Zhu, X., Zambaldo, C., Liu, R., Liu, T., Lyu, X., Du, J., and Xuan, W. (2017). Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. Nature chemical biology 13, 845.

Magliery, T.J., Anderson, J.C., and Schultz, P.G. (2001). Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in Escherichia coli1. Journal of molecular biology 307, 755-769.

Makarewich, C.A., and Olson, E.N. (2017). Mining for micropeptides. Trends in cell biology 27, 685-696.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science 298, 1912-1934.

Martin, B.R., Giepmans, B.N., Adams, S.R., and Tsien, R.Y. (2005). Mammalian cell–based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. Nature biotechnology 23, 1308.

Masuda, T., Petrov, A.N., Iizuka, R., Funatsu, T., Puglisi, J.D., and Uemura, S. (2012). Initiation factor 2, tRNA, and 50S subunits cooperatively stabilize mRNAs on the ribosome during initiation. Proceedings of the National Academy of Sciences.

McMahon, A., Supatto, W., Fraser, S.E., and Stathopoulos, A. (2008). Dynamic analyses of Drosophila gastrulation provide insights into collective cell migration. Science 322, 1546-1550.

Mehl, R.A., Anderson, J.C., Santoro, S.W., Wang, L., Martin, A.B., King, D.S., Horn, D.M., and Schultz, P.G. (2003). Generation of a bacterium with a 21 amino acid genetic code. Journal of the American Chemical Society 125, 935-939.

Melnikov, S., Ben-Shem, A., De Loubresse, N.G., Jenner, L., Yusupova, G., and Yusupov, M. (2012). One core, two shells: bacterial and eukaryotic ribosomes. Nature structural & molecular biology 19, 560.

Merrifield, R.B. (1963). Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. Journal of the American Chemical Society 85, 2149-2154.

Michaels, M.L., Kim, C.W., Matthews, D.A., and Miller, J.H. (1990). Escherichia coli thymidylate synthase: amino acid substitutions by suppression of amber nonsense mutations. Proceedings of the National Academy of Sciences 87, 3957-3961.

Milón, P., Maracci, C., Filonava, L., Gualerzi, C.O., and Rodnina, M.V. (2012). Real-time assembly landscape of bacterial 30S translation initiation complex. Nature structural & molecular biology 19, 609.

Moore, B., Persson, B.C., Nelson, C.C., Gesteland, R.F., and Atkins, J.F. (2000). Quadruplet codons: implications for code expansion and the specification of translation step size1. Journal of molecular biology 298, 195-209.

Mühlhausen, S., Schmitt, H.D., Pan, K.-T., Plessmann, U., Urlaub, H., Hurst, L.D., and Kollmar, M. (2018). Endogenous Stochastic Decoding of the CUG Codon by Competing Ser- and Leu-tRNAs in Ascoidea asiatica. Current Biology.

Muhs, M., Hilal, T., Mielke, T., Skabkin, M.A., Sanbonmatsu, K.Y., Pestova, T.V., and Spahn, C.M. (2015). Cryo-EM of ribosomal 80S complexes with termination factors reveals the translocated cricket paralysis virus IRES. Molecular cell 57, 422-432.

Mujumdar, R.B., Ernst, L.A., Mujumdar, S.R., Lewis, C.J., and Waggoner, A.S. (1993). Cyanine dye labelling reagents: sulfoindocyanine succinimidyl esters. Bioconjugate chemistry 4, 105-111.

Mukai, T., Hayashi, A., Iraha, F., Sato, A., Ohtake, K., Yokoyama, S., and Sakamoto, K. (2010a). Codon reassignment in the Escherichia coli genetic code. Nucleic acids research 38, 8188-8195.

Mukai, T., Kobayashi, T., Hino, N., Yanagisawa, T., Sakamoto, K., and Yokoyama, S. (2008). Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. Biochemical and biophysical research communications 371, 818-822.

Mukai, T., Wakiyama, M., Sakamoto, K., and Yokoyama, S. (2010b). Genetic encoding of non-natural amino acids in Drosophila melanogaster Schneider 2 cells. Protein Science 19, 440-448.

Mukherjee, M., Sabir, S., O'Regan, L., Sampson, J., Richards, M.W., Huguenin-Dezot, N., Ault, J.R., Chin, J.W., Zhuravleva, A., and Fry, A.M. (2018). Mitotic phosphorylation regulates Hsp72 spindle localization by uncoupling ATP binding from substrate release. Sci Signal 11, eaao2464.

Nakamura, Y., Ito, K., Matsumura, K., Kawazu, Y., and Ebihara, K. (1995). Regulation of translation termination: conserved structural motifs in bacterial and eukaryotic polypeptide release factors. Biochemistry and cell biology 73, 1113-1122.

Namy, O., Rousset, J.-P., Napthine, S., and Brierley, I. (2004). Reprogrammed genetic decoding in cellular gene expression. Molecular cell 13, 157-168.

Neumann, H., Hancock, S.M., Buning, R., Routh, A., Chapman, L., Somers, J., Owen-Hughes, T., van Noort, J., Rhodes, D., and Chin, J.W. (2009). A method for genetically installing site-

specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. Molecular cell 36, 153-163.

Neumann, H., Hazen, J.L., Weinstein, J., Mehl, R.A., and Chin, J.W. (2008a). Genetically encoding protein oxidative damage. Journal of the American Chemical Society 130, 4028-4033.

Neumann, H., Peak-Chew, S.Y., and Chin, J.W. (2008b). Genetically encoding Nε-acetyllysine in recombinant proteins. Nature chemical biology 4, 232-234.

Neumann, H., Wang, K., Davis, L., Garcia-Alai, M., and Chin, J.W. (2010). Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. Nature 464, 441.

Newberry, K.J., Hou, Y.M., and Perona, J.J. (2002). Structural origins of amino acid selection without editing by cysteinyl-tRNA synthetase. The EMBO journal 21, 2778-2787.

Nguyen, D.P., Alai, M.M.G., Virdee, S., and Chin, J.W. (2010). Genetically directing ε-N, N-dimethyl-L-lysine in recombinant histones. Chemistry & biology 17, 1072-1076.

Nguyen, D.P., Elliott, T., Holt, M., Muir, T.W., and Chin, J.W. (2011). Genetically encoded 1, 2-aminothiols facilitate rapid and site-specific protein labelling via a bio-orthogonal cyanobenzothiazole condensation. Journal of the American Chemical Society 133, 11418-11421.

Nguyen, D.P., Garcia Alai, M.M., Kapadnis, P.B., Neumann, H., and Chin, J.W. (2009a). Genetically Encoding N ϵ-Methyl-l-lysine in Recombinant Histones. Journal of the American Chemical Society 131, 14194-14195.

Nguyen, D.P., Lusic, H., Neumann, H., Kapadnis, P.B., Deiters, A., and Chin, J.W. (2009b). Genetic encoding and labelling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNACUA pair and click chemistry. Journal of the American Chemical Society 131, 8720-8721.

Nguyen, D.P., Mahesh, M., Elsässer, S.J., Hancock, S.M., Uttamapinant, C., and Chin, J.W. (2014). Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. Journal of the American Chemical Society 136, 2240-2243.

Nieuwenhuizen, R.P., Lidke, K.A., Bates, M., Puig, D.L., Grünwald, D., Stallinga, S., and Rieger, B. (2013). Measuring image resolution in optical nanoscopy. Nature methods 10, 557.

Nikić, I., Estrada Girona, G., Kang, J.H., Paci, G., Mikhaleva, S., Koehler, C., Shymanska, N.V., Ventura Santos, C., Spitz, D., and Lemke, E.A. (2016). Debugging Eukaryotic Genetic Code Expansion for Site-Specific Click-PAINT Super-Resolution Microscopy. Angewandte Chemie International Edition 55, 16172-16176.

Nirenberg, M.W., and Matthaei, J.H. (1961). The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proceedings of the National Academy of Sciences 47, 1588-1602.

Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C., and Schultz, P.G. (1989). A general method for site-specific incorporation of unnatural amino acids into proteins. Science 244, 182-188.

Nowak, M.W., Kearney, P.C., Saks, M., Labarca, C., Silverman, S., Zhong, W., Thorson, J., Abelson, J., and Davidson, N. (1995). Nicotinic receptor binding site probed with unnatural amino acid incorporation in intact cells. Science 268, 439-442.

Obata, T., and Shiraiwa, Y. (2005). A novel eukaryotic selenoprotein in the haptophyte alga Emiliania huxleyi. Journal of Biological Chemistry.

Oller-Salvia, B., Kym, G., & Chin, J. W. (2018). Rapid and Efficient Generation of Stable Antibody–Drug Conjugates via an Encoded Cyclopropene and an Inverse-Electron-Demand Diels–Alder Reaction. Angewandte Chemie International Edition, 57(11), 2831-2834.

Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127, 635-648.

Ostrov, N., Landon, M., Guell, M., Kuznetsov, G., Teramoto, J., Cervantes, N., Zhou, M., Singh, K., Napolitano, M.G., and Moosburner, M. (2016). Design, synthesis, and testing toward a 57-codon genome. Science 353, 819-822.

Page, M.F., Carr, B., Anders, K.R., Grimson, A., and Anderson, P. (1999). SMG-2 is a phosphorylated protein required for mRNA surveillance in Caenorhabditis elegans and related to Upf1p of yeast. Molecular and cellular biology 19, 5943-5951.

Panchuk-Voloshina, N., Haugland, R.P., Bishop-Stewart, J., Bhalgat, M.K., Millard, P.J., Mao, F., Leung, W.-Y., and Haugland, R.P. (1999). Alexa dyes, a series of new fluorescent dyes that yield exceptionally bright, photostable conjugates. Journal of Histochemistry & Cytochemistry 47, 1179-1188.

Park, H.-S., Hohn, M.J., Umehara, T., Guo, L.-T., Osborne, E.M., Benner, J., Noren, C.J., Rinehart, J., and Söll, D. (2011). Expanding the genetic code of Escherichia coli with phosphoserine. Science 333, 1151-1154.

Parrish, A.R., She, X., Xiang, Z., Coin, I., Shen, Z., Briggs, S.P., Dillin, A., and Wang, L. (2012). Expanding the genetic code of Caenorhabditis elegans using bacterial aminoacyl-tRNA synthetase/tRNA pairs. ACS chemical biology 7, 1292-1302.

Pattabiraman, V.R., and Bode, J.W. (2011). Rethinking amide bond synthesis. Nature 480, 471.

Patterson, G.H., and Lippincott-Schwartz, J. (2002). A photoactivatable GFP for selective photolabelling of proteins and cells. Science 297, 1873-1877.

Pawson, T., and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. science 300, 445-452.

Pearlman, S.M., Serber, Z., and Ferrell Jr, J.E. (2011). A mechanism for the evolution of phosphorylation sites. Cell 147, 934-946.

Peng, T., and Hang, H.C. (2016). Site-specific bioorthogonal labelling for fluorescence imaging of intracellular proteins in living cells. Journal of the American Chemical Society 138, 14423-14433.

Pestova, T.V., Lorsch, J.R., and Hellen, C.U. (2007). The mechanism of translation initiation in eukaryotes. Cold Spring Harbor Monograph Series 48, 87.

Pinsent, J. (1954). The need for selenite and molybdate in the formation of formic dehydrogenase by members of the coli-aerogenes group of bacteria. Biochemical Journal 57, 10.

Pisarev, A.V., Skabkin, M.A., Pisareva, V.P., Skabkina, O.V., Rakotondrafara, A.M., Hentze, M.W., Hellen, C.U., and Pestova, T.V. (2010). The role of ABCE1 in eukaryotic posttermination ribosomal recycling. Molecular cell 37, 196-210.

Plass, T., Milles, S., Koehler, C., Schultz, C., and Lemke, E.A. (2011). Genetically encoded copper-free click chemistry. Angewandte Chemie International Edition 50, 3878-3881.

Plass, T., Milles, S., Koehler, C., Szymański, J., Mueller, R., Wießler, M., Schultz, C., and Lemke, E.A. (2012). Amino acids for Diels–Alder reactions in living cells. Angewandte Chemie 124, 4242-4246.

Polycarpo, C., Ambrogelly, A., Bérubé, A., Winbush, S.M., McCloskey, J.A., Crain, P.F., Wood, J.L., and Söll, D. (2004). An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. Proceedings of the National Academy of Sciences 101, 12450-12454.

Polycarpo, C.R., Herring, S., Bérubé, A., Wood, J.L., Söll, D., and Ambrogelly, A. (2006). Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. FEBS letters 580, 6695-6700.

Prasher, D.C., Eckenrode, V.K., Ward, W.W., Prendergast, F.G., and Cormier, M.J. (1992). Primary structure of the Aequorea victoria green-fluorescent protein. Gene 111, 229-233.

Prat, L., Heinemann, I.U., Aerni, H.R., Rinehart, J., O'Donoghue, P., and Söll, D. (2012). Carbon source-dependent expansion of the genetic code in bacteria. Proceedings of the National Academy of Sciences 109, 21070-21075.

Rackham, O., and Chin, J.W. (2005). A network of orthogonal ribosome· mRNA pairs. Nature chemical biology 1, 159.

Ramil, C.P., Dong, M., An, P., Lewandowski, T.M., Yu, Z., Miller, L.J., and Lin, Q. (2017). Spirohexene-tetrazine ligation enables bioorthogonal labelling of class BG protein-coupled receptors in live cells. Journal of the American Chemical Society 139, 13376-13386.

Rasmussen, L.C., Laursen, B.S., Mortensen, K.K., and Sperling-Petersen, H.U. (2009). Initiator tRNAs in bacteria and eukaryotes. eLS.

Reeves, M., and Hoffmann, P. (2009). The human selenoproteome: recent insights into functions and regulation. Cellular and molecular life sciences 66, 2457-2478.

Richmond, M. (1962). The effect of amino acid analogues on growth and protein synthesis in microorganisms. Bacteriological reviews 26, 398.

Riddle, D.L., and Carbon, J. (1973). Frameshift suppression: a nucleotide addition in the anticodon of a glycine transfer RNA. Nature New Biology 242, 230.

Roberts-Galbraith, R.H., Ohi, M.D., Ballif, B.A., Chen, J.-S., McLeod, I., McDonald, W.H., Gygi, S.P., Yates III, J.R., and Gould, K.L. (2010). Dephosphorylation of F-BAR protein Cdc15 modulates its conformation and stimulates its scaffolding activity at the cell division site. Molecular cell 39, 86-99.

Rogerson, D.T., Sachdeva, A., Wang, K., Haq, T., Kazlauskaite, A., Hancock, S.M., Huguenin-Dezot, N., Muqit, M.M., Fry, A.M., Bayliss, R., et al. (2015). Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. Nat Chem Biol 11, 496-503.

Rould, M.A., Perona, J.J., Soll, D., and Steitz, T.A. (1989). Structure of E. coli glutaminyl-tRNA synthetase complexed with tRNA (Gln) and ATP at 2.8 A resolution. Science 246, 1135-1142.

Roy, B., Leszyk, J.D., Mangus, D.A., and Jacobson, A. (2015). Nonsense suppression by near-cognate tRNAs employs alternative base pairing at codon positions 1 and 3. Proceedings of the National Academy of Sciences, 201424127.

Rust, M.J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nature methods 3, 793.

Rydén, S., and Isaksson, L. (1984). A temperature-sensitive mutant of Escherichia coli that shows enhanced misreading of UAG/A and increased efficiency for tRNA nonsense suppressors. Molecular and General Genetics MGG 193, 38-45.

Sakamoto, K., Hayashi, A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., Kobayashi, T., Kitabatake, M., Takio, K., and Saito, K. (2002). Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. Nucleic Acids Res 30, 4692-4699.

Sauerwald, A., Zhu, W., Major, T.A., Roy, H., Palioura, S., Jahn, D., Whitman, W.B., Yates, J.R., Ibba, M., and Söll, D. (2005). RNA-dependent cysteine biosynthesis in archaea. Science 307, 1969-1972.

Schmidt, M.J., Borbas, J., Drescher, M., and Summerer, D. (2014). A genetically encoded spin label for electron paramagnetic resonance distance measurements. Journal of the American Chemical Society 136, 1238-1241.

Schmidt, M.J., and Summerer, D. (2013). Red-Light-Controlled Protein–RNA Crosslinking with a Genetically Encoded Furan. Angewandte Chemie International Edition 52, 4690-4693.

Schmied, W.H., Elsasser, S.J., Uttamapinant, C., and Chin, J.W. (2014). Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1. J Am Chem Soc 136, 15577-15583.

Schnell, U., Dijk, F., Sjollema, K.A., and Giepmans, B.N. (2012). Immunolabelling artifacts and the need for live-cell imaging. Nature methods 9, 152.

Schrader, J.M., Chapman, S.J., and Uhlenbeck, O.C. (2011). Tuning the affinity of aminoacyl-tRNA to elongation factor Tu for optimal decoding. Proceedings of the National Academy of Sciences.

Schrader, J.M., and Uhlenbeck, O.C. (2011). Is the sequence-specific binding of aminoacyl-tRNAs by EF-Tu universal among bacteria? Nucleic acids research 39, 9746-9758.

Senger, B., Auxilien, S., Englisch, U., Cramer, F., and Fasiolo, F. (1997). The modified wobble base inosine in yeast tRNAIle is a positive determinant for aminoacylation by isoleucyl-tRNA synthetase. Biochemistry 36, 8269-8275.

Serwa, R., Wilkening, I., Del Signore, G., Mühlberg, M., Claußnitzer, I., Weise, C., Gerrits, M., and Hackenberger, C.P. (2009). Chemoselective Staudinger-phosphite reaction of azides for the phosphorylation of proteins. Angewandte Chemie International Edition 48, 8234-8239.

Sharp, S.J., Schaack, J., Cooley, L., Burke, D.J., and Soil, D. (1985). Structure and transcription of eukaryotic tRNA gene. Critical Reviews In Biochemistry 19, 107-144.

Shoemaker, C.J., and Green, R. (2011). Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. Proceedings of the National Academy of Sciences 108, E1392-E1398.

Slavoff, S.A., Mitchell, A.J., Schwaid, A.G., Cabili, M.N., Ma, J., Levin, J.Z., Karger, A.D., Budnik, B.A., Rinn, J.L., and Saghatelian, A. (2013). Peptidomic discovery of short open reading frame–encoded peptides in human cells. Nature chemical biology 9, 59.

Söil, D. (1988). Enter a new amino acid. Nature 331, 662.

Söll, D. (1990). The accuracy of aminoacylation—ensuring the fidelity of the genetic code. Experientia 46, 1089-1096.

Srinivasan, G., James, C.M., and Krzycki, J.A. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science 296, 1459-1462.

St-Pierre, F., Marshall, J.D., Yang, Y., Gong, Y., Schnitzer, M.J., and Lin, M.Z. (2014). High-fidelity optical reporting of neuronal electrical activity with an ultrafast fluorescent voltage sensor. Nature neuroscience 17, 884.

Stadler, C., Rexhepaj, E., Singan, V.R., Murphy, R.F., Pepperkok, R., Uhlén, M., Simpson, J.C., and Lundberg, E. (2013). Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells. Nature methods 10, 315.

Stark, H., Rodnina, M.V., Rinke-Appel, J., Brimacombe, R., Wintermeyer, W., and van Heel, M. (1997). Visualization of elongation factor Tu on the Escherichia coli ribosome. Nature 389, 403.

Steeg, P.S., Palmieri, D., Ouatas, T., and Salerno, M. (2003). Histidine kinases and histidine phosphorylated proteins in mammalian cell biology, signal transduction and cancer. Cancer letters 190, 1-12.

Steer, B.A., and Schimmel, P. (1999). Major anticodon-binding region missing from an archaebacterial tRNA synthetase. Journal of Biological Chemistry 274, 35601-35606.

Stemmer, W.P. (1994). Rapid evolution of a protein in vitro by DNA shuffling. Nature 370, 389.

Sulston, J.E., and Horvitz, H.R. (1977). Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. Developmental biology 56, 110-156.

Sykes, B.D., Weingarten, H.I., and Schlesinger, M.J. (1974). Fluorotyrosine alkaline phosphatase from Escherichia coli: preparation, properties, and fluorine-19 nuclear magnetic resonance spectrum. Proceedings of the National Academy of Sciences 71, 469-473.

Szymborska, A., de Marco, A., Daigle, N., Cordes, V.C., Briggs, J.A., and Ellenberg, J. (2013). Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. Science 341, 655-658.

Taylor, D., Unbehaun, A., Li, W., Das, S., Lei, J., Liao, H.Y., Grassucci, R.A., Pestova, T.V., and Frank, J. (2012). Cryo-EM structure of the mammalian eukaryotic release factor eRF1–eRF3-associated termination complex. Proceedings of the National Academy of Sciences, 201216730.

Taylor, D.J., Frank, J., and Kinzy, T.G. (2007). Structure and function of the eukaryotic ribosome and elongation factors. COLD SPRING HARBOR MONOGRAPH SERIES 48, 59.

Tharp, J.M., Wang, Y.-S., Lee, Y.-J., Yang, Y., and Liu, W.R. (2014). Genetic incorporation of seven ortho-substituted phenylalanine derivatives. ACS chemical biology 9, 884-890.

Torbeev, V.Y., and Kent, S.B. (2007). Convergent chemical synthesis and crystal structure of a 203 amino acid “covalent dimer” HIV-1 protease enzyme molecule. Angewandte Chemie International Edition 46, 1667-1670.

Tornøe, C.W., Christensen, C., and Meldal, M. (2002). Peptidotriazoles on solid phase:[1, 2, 3]-triazoles by regiospecific copper (I)-catalyzed 1, 3-dipolar cycloadditions of terminal alkynes to azides. The Journal of organic chemistry 67, 3057-3064.

Trezeguet, V., Edwards, H., and Schimmel, P. (1991). A single base pair dominates over the novel identity of an Escherichia coli tyrosine tRNA in Saccharomyces cerevisiae. Molecular and cellular biology 11, 2744-2751.

Trichas, G., Smith, A.M., White, N., Wilkins, V., Watanabe, T., Moore, A., Joyce, B., Sugnaseelan, J., Rodriguez, T.A., and Kay, D. (2012). Multi-cellular rosettes in the mouse visceral endoderm facilitate the ordered migration of anterior visceral endoderm cells. PLoS biology 10, e1001256.

Tropberger, P., Pott, S., Keller, C., Kamieniarz-Gdula, K., Caron, M., Richter, F., Li, G., Mittler, G., Liu, E.T., and Bühler, M. (2013). Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. Cell 152, 859-872.

Tsien, R.Y. (1998). The green fluorescent protein (Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA).

Tsugita, A., and Fraenkel-Conrat, H. (1960). The amino acid composition and C-terminal sequence of a chemically evoked mutant of TMV. Proceedings of the National Academy of Sciences 46, 636-642.

Tujebajeva, R.M., Copeland, P.R., Xu, X.M., Carlson, B.A., Harney, J.W., Driscoll, D.M., Hatfield, D.L., and Berry, M.J. (2000). Decoding apparatus for eukaryotic selenocysteine insertion. EMBO reports 1, 158-163.

Tuley, A., Wang, Y.-S., Fang, X., Kurra, Y., Rezenom, Y.H., and Liu, W.R. (2014). The genetic incorporation of thirteen novel non-canonical amino acids. Chemical Communications 50, 2673-2675.

Uttamapinant, C., Howe, J.D., Lang, K., Beránek, V.c., Davis, L., Mahesh, M., Barry, N.P., and Chin, J.W. (2015). Genetic code expansion enables live-cell and super-resolution imaging of site-specifically labeled cellular proteins. Journal of the American Chemical Society 137, 4602-4605.

Uttamapinant, C., White, K.A., Baruah, H., Thompson, S., Fernández-Suárez, M., Puthenveetil, S., and Ting, A.Y. (2010). A fluorophore ligase for site-specific protein labelling inside living cells. Proceedings of the National Academy of Sciences 107, 10914-10919.

Villén, J., Beausoleil, S.A., Gerber, S.A., and Gygi, S.P. (2007). Large-scale phosphorylation analysis of mouse liver. Proceedings of the National Academy of Sciences 104, 1488-1493.

Virdee, S., Kapadnis, P.B., Elliott, T., Lang, K., Madrzak, J., Nguyen, D.P., Riechmann, L., and Chin, J.W. (2011). Traceless and site-specific ubiquitination of recombinant proteins. Journal of the American Chemical Society 133, 10708-10711.

Virdee, S., Ye, Y., Nguyen, D.P., Komander, D., and Chin, J.W. (2010). Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. Nature chemical biology 6, 750.

Vogel, Z., Zamir, A., and Elson, D. (1969). Possible involvement of peptidyl transferase in the termination step of protein biosynthesis. Biochemistry 8, 5161-5168.

Walsh, C. (2006). Posttranslational modification of proteins: expanding nature's inventory (Roberts and Company Publishers).

Walsh, C.T., Garneau-Tsodikova, S., and Gatto Jr, G.J. (2005). Protein posttranslational modifications: the chemistry of proteome diversifications. Angewandte Chemie International Edition 44, 7342-7372.

Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R., and Church, G.M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. Nature 460, 894.

Wang, K., Fredens, J., Brunner, S.F., Kim, S.H., Chia, T., and Chin, J.W. (2016). Defining synonymous codon compression schemes by genome recoding. Nature 539, 59.

Wang, K., Neumann, H., Peak-Chew, S.Y., and Chin, J.W. (2007a). Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. Nature biotechnology 25, 770.

Wang, L., Magliery, T.J., Liu, D.R., and Schultz, P.G. (2000). A New Functional Suppressor tRNA/Aminoacyl− tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins. Journal of the American Chemical Society 122, 5010-5011.

Wang, Q., Chan, T.R., Hilgraf, R., Fokin, V.V., Sharpless, K.B., and Finn, M. (2003). Bioconjugation by copper (I)-catalyzed azide-alkyne [3+ 2] cycloaddition. Journal of the american chemical society 125, 3192-3193.

Wang, S., Moffitt, J.R., Dempsey, G.T., Xie, X.S., and Zhuang, X. (2014). Characterization and development of photoactivatable fluorescent proteins for single-molecule–based superresolution imaging. Proceedings of the National Academy of Sciences, 201406593.

Wang, W., Takimoto, J.K., Louie, G.V., Baiga, T.J., Noel, J.P., Lee, K.-F., Slesinger, P.A., and Wang, L. (2007b). Genetically encoding unnatural amino acids for cellular and neuronal studies. Nature neuroscience 10, 1063.

Wang, Y.-S., Fang, X., Chen, H.-Y., Wu, B., Wang, Z.U., Hilty, C., and Liu, W.R. (2012a). Genetic incorporation of twelve meta-substituted phenylalanine derivatives using a single pyrrolysyl-tRNA synthetase mutant. ACS chemical biology 8, 405-415.

Wang, Y.-S., Fang, X., Wallace, A.L., Wu, B., and Liu, W.R. (2012b). A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum. Journal of the American Chemical Society 134, 2950-2953.

Wang, Y.-S., Wu, B., Wang, Z., Huang, Y., Wan, W., Russell, W.K., Pai, P.-J., Moe, Y.N., Russell, D.H., and Liu, W.R. (2010). A genetically encoded photocaged Nε-methyl-l-lysine. Molecular BioSystems 6, 1557-1560.

Wang, Z.A., Zeng, Y., Kurra, Y., Wang, X., Tharp, J.M., Vatansever, E.C., Hsu, W.W., Dai, S., Fang, X., and Liu, W.R. (2017). A Genetically Encoded Allysine for the Synthesis of Proteins with Site-Specific Lysine Dimethylation. Angewandte Chemie International Edition 56, 212-216.

Washburn, T., and O'Tousa, J. (1992). Nonsense suppression of the major rhodopsin gene of Drosophila. Genetics 130, 585-595.

Whelihan, E.F., and Schimmel, P. (1997). Rescuing an essential enzyme–RNA complex with a non-essential appended domain. The EMBO Journal 16, 2968-2974.

Wiedenmann, J., Ivanchenko, S., Oswald, F., Schmitt, F., Röcker, C., Salih, A., Spindler, K.-D., and Nienhaus, G.U. (2004). EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. Proceedings of the National Academy of Sciences 101, 15905-15910.

Willis, J.C.W., and Chin, J.W. (2018). Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. Nat Chem 10, 831-837.

Wittmann, H.-G. (1961). Studies on the nucleic acid-protein correlation in tobacco mosaic virus. Paper presented at: Proceedings of the 5th International Congress of Biochemistry, Moscow 10.

Wolbers, F., ter Braak, P., Le Gac, S., Luttge, R., Andersson, H., Vermes, I., and van den Berg, A. (2006). Viability study of HL60 cells in contact with commonly used microchip materials. Electrophoresis 27, 5073-5080.

Wu, N., Deiters, A., Cropp, T.A., King, D., and Schultz, P.G. (2004). A genetically encoded photocaged amino acid. J Am Chem Soc 126, 14306-14307.

Xiao, H., Chatterjee, A., Choi, S.h., Bajjuri, K.M., Sinha, S.C., and Schultz, P.G. (2013). Genetic incorporation of multiple unnatural amino acids into proteins in mammalian cells. Angew Chem, Int Ed 52, 14080-14083.

Xiao, H., Peters, F.B., Yang, P.-Y., Reed, S., Chittuluru, J.R., and Schultz, P.G. (2014). Genetic incorporation of histidine derivatives using an engineered pyrrolysyl-tRNA synthetase. ACS chemical biology 9, 1092-1096.

Xie, J., Supekova, L., and Schultz, P.G. (2007). A genetically encoded metabolically stable analogue of phosphotyrosine in Escherichia coli. ACS chemical biology 2, 474-478.

Yada, M., Hatakeyama, S., Kamura, T., Nishiyama, M., Tsunematsu, R., Imaki, H., Ishida, N., Okumura, F., Nakayama, K., and Nakayama, K.I. (2004). Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7. The EMBO journal 23, 2116-2125.

Yadavalli, S.S., and Ibba, M. (2012). Quality control in aminoacyl-tRNA synthesis: Its role in translational fidelity. In Advances in protein chemistry and structural biology (Elsevier), pp. 1-43.

Yamane, T., and Hopfield, J. (1977). Experimental evidence for kinetic proofreading in the aminoacylation of tRNA by synthetase. Proceedings of the National Academy of Sciences 74, 2246-2250.

Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., and Yokoyama, S. (2008). Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode Nε-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. Chemistry & biology 15, 1187-1197.

Yang, A., Ha, S., Ahn, J., Kim, R., Kim, S., Lee, Y., Kim, J., Söll, D., Lee, H.-Y., and Park, H.-S. (2016). A chemical biology route to site-specific authentic protein modifications. Science 354, 623-626.

Yourno, J., and Kohno, T. (1972). Externally Suppressible Proline Quadruplet CCCUU. Science 175, 650-652.

Yuan, J., Palioura, S., Salazar, J.C., Su, D., O'Donoghue, P., Hohn, M.J., Cardoso, A.M., Whitman, W.B., and Söll, D. (2006). RNA-dependent conversion of phosphoserine forms

selenocysteine in eukaryotes and archaea. Proceedings of the National Academy of Sciences 103, 18923-18927.

Zhang, J., Yang, P.L., and Gray, N.S. (2009). Targeting cancer with small molecule kinase inhibitors. Nature Reviews Cancer 9, 28.

Zhang, M., Lin, S., Song, X., Liu, J., Fu, Y., Ge, X., Fu, X., Chang, Z., and Chen, P.R. (2011). A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. Nature chemical biology 7, 671-677.

Zhang, M.S., Brunner, S.F., Huguenin-Dezot, N., Liang, A.D., Schmied, W.H., Rogerson, D.T., and Chin, J.W. (2017). Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. Nat Methods 14, 729.

Zhang, Y., Baranov, P.V., Atkins, J.F., and Gladyshev, V.N. (2005). Pyrrolysine and selenocysteine use dissimilar decoding strategies. Journal of Biological Chemistry.

Zhang, Y., Romero, H., Salinas, G., and Gladyshev, V.N. (2006). Dynamic evolution of selenocysteine utilization in bacteria: a balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues. Genome biology 7, R94.

Zhang, Z., Alfonta, L., Tian, F., Bursulaya, B., Uryu, S., King, D.S., and Schultz, P.G. (2004). Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. Proceedings of the National Academy of Sciences 101, 8882-8887.

Zhang, Z., Wang, L., Brock, A., and Schultz, P.G. (2002). The selective incorporation of alkenes into proteins in Escherichia coli. Angewandte Chemie International Edition 41, 2840-2842.

Zheng, Y., Addy, P.S., Mukherjee, R., and Chatterjee, A. (2017a). Defining the current scope and limitations of dual noncanonical amino acid mutagenesis in mammalian cells. Chem Sci 8, 7211-7217.

Zheng, Y., Mukherjee, R., Chin, M.A., Igo, P., Gilgenast, M.J., and Chatterjee, A. (2017b). Expanding the Scope of Single-and Double-Noncanonical Amino Acid Mutagenesis in

Mammalian Cells Using Orthogonal Polyspecific Leucyl-tRNA Synthetases. Biochemistry 57, 441-445.

Zimmer, M. (2002). Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. Chemical reviews 102, 759-782.

Zinoni, F., Birkmann, A., Stadtman, T.C., and Böck, A. (1986). Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from Escherichia coli. Proceedings of the National Academy of Sciences 83, 4650-465